AN IMPROVED ALGORITHM FOR THE DETERMINATION OF THE SYSTEM PARAMETERS OF A VISUAL BINARY BY LEAST SQUARES

By

YU-LIN XU

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1988

To my mother and late father

ACKNOWLEDGEMENTS

The author acknowledges his heartfelt gratitude to Dr. Heinrich K. Eichhorn, his research advisor, for proposing this subject, the considerate guidance and encouragement throughout his research, and for the patience in reading and correcting the manuscript. The author has benefited in many ways as Dr. Eichhorn's student.

The author is also grateful to Drs. Kwan-Yu Chen, Haywood C. Smith, Frank Bradshaw Wood and Philip Bacon for having served as members of his supervisory committee and for helpful discussions, timely suggestions and the careful review of this dissertation. Likewise, his deep appreciation goes to Drs. W. D. Heintz and H. A. Macalister for having provided data used in his dissertation.

The author is especially grateful to Drs. Jerry L. Weinberg and Ru-Tsan Wang in the Space Astronomy Laboratory for their considerate encouragement and support. Without their support, the fulfillment of this research could not be possible.

It is a great pleasure to acknowledge that all the calculations were performed on the Vax in the Space Astronomy Laboratory.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

AN IMPROVED ALGORITHM FOR
THE DETERMINATION OF THE SYSTEM PARAMETERS
OF A VISUAL BINARY BY LEAST SQUARES

By

YU-LIN XU

April 1988

Chairman: Dr. Heinrich K. Eichhorn
Co-Chairman: Dr. Kwan-Yu Chen
Major Department: Astronomy

The problem of computing the orbit of a visual binary from a set of observed positions is reconsidered. It is a least squares adjustment problem, if the observational errors follow a bias-free multivariate Gaussian distribution and the covariance matrix of the observations is assumed to be known.

The condition equations are constructed to satisfy both the conic section equation and the area theorem, which are nonlinear in both the observations and the adjustment parameters. The traditional least squares algorithm, which employs condition equations that are solved with respect to the uncorrelated observations and either linear in the adjustment parameters or linearized by developing them in Taylor series by first-order approximation, is inadequate in

our orbit problem. D. C. Brown proposed an algorithm solving a more general least squares adjustment problem in which the scalar residual function, however, is still constructed by first-order approximation. Not long ago, a completely general solution was published by W. H. Jefferys, who proposed a rigorous adjustment algorithm for models in which the observations appear nonlinearly in the condition equations and may be correlated, and in which construction of the normal equations and the residual function involves no approximation. This method was successfully applied in our problem.

The normal equations were first solved by Newton's scheme. Practical examples show that this converges fast if the observational errors are sufficiently small and the initial approximate solution is sufficiently accurate, and that it fails otherwise. Newton's method was modified to yield a definitive solution in the case the normal approach fails, by combination with the method of steepest descent and other sophisticated algorithms. Practical examples show that the modified Newton scheme can always lead to a final solution.

The weighting of observations, the orthogonal parameters and the "efficiency" of a set of adjustment parameters are also considered. The definition of "efficiency" is revised.

CHAPTER I
INTRODUCTION

The problem of computing the orbit of a visual binary from a set of observed positions is by no means new. A great variety of methods has been proposed. As is well known, only a few of these suffice to cover the practical contingencies, and the majority fails to handle the input data efficiently and properly.

For the visual binary case, the determination of an orbit normally requires a large number of observations. All measures of position angles and separations are, as are all observations, affected by observational errors. For the purpose of our work, these errors are assumed to follow a bias-free multivariate Gaussian distribution. Under this assumption, orbit-computing is a least squares adjustment problem, in which the condition equations are nonlinear in both the observations and the adjustment parameters. The condition equations must incorporate all relationships that exist between observations and the orbital parameters,*

---

*Usually, the term "orbital elements" is used. We prefer "orbital parameters" instead. Strictly speaking, the orbital elements are the constants of integration in the two-body problem and, therefore, do not include the masses of the components.

that is, must state both the area theorem, which follows from Kepler's equation, and the condition that the projected orbit is a conic section. H. Eichhorn (1985) has suggested a new form for the construction of a complete set of condition equations for this very problem.

The traditional least squares algorithm, which is based on condition equations linearized with respect to the observational errors, will not lead to those orbital parameters which minimize the sum of the squares of observational errors, because linearization, in this case, is too crude an approximation. In some earlier papers, H. Eichhorn and W. G. Clary (1974) proposed a least squares solution algorithm, which takes into account the second (in addition to the first) order derivatives in the adjustment residuals (observational errors) and the corrections to the initially available approximation to the adjustment parameters. A completely general solution was published by W. H. Jefferys (1980, 1981), who proposed a rigorous adjustment algorithm for models in which the observations appear nonlinearly in the condition equations. In addition, there may be nonlinear constraints** among model parameters, and the observations may be correlated. In practice, the method is nearly as simple to apply as the classical method of least

---

**We use the term "constraints" for condition equations which do not contain any observations explicitly.

squares, for it does not require calculation of any derivatives of higher than the first order.

In this paper, we present an approach to solve the orbit problem by Jefferys' method, in which both the area theorem and the conic section equation assume the function of the condition equations.

CHAPTER II
REVIEW AND REMARKS

Review of the Methods for Orbit-Computation

Every complete observation of a double star supplies us with three data: the time of observation, the position angle of the secondary with respect to the primary, and the angular distance (separation) between the two stars. The problem of computing the so-called orbital elements (in this paper, called orbital parameters) of a visual binary from a set of observations superficially appears analogous to the case of orbits in the planetary system, yet in practice there is little resemblance between the problems. The problem of determining the orbit of a body in the solar system is complicated by the motion of the observer who shares the motion of the Earth, so that, unlike in the case of a binary, the apparent path is not merely a projection of the orbit in space onto the celestial sphere.

In the case of a binary, the path observed is the projection of the motion of the secondary round the primary onto a plane perpendicular to the line of sight. The apparent orbit (i.e., the observed path of the secondary about the primary) is therefore not a mere scale drawing of the true orbit in space. The primary may be situated at any

point within the ellipse described by the secondary and, of course, does not necessarily occupy either a focus or the center.

The problem of deriving "an orbit" (meaning a set of estimates of the orbital parameters) from the observations was first solved by F. Savary in 1827. In 1829, J. F. Encke quickly followed with a different solution method which was somewhat better adapted to what were then the needs of the practical astronomer. Theoretically, the methods of Savary and Encke are excellent. But their methods utilize only four complete pairs of measures (angle and distance) instead of all the available data and closely emulate the treatment of planetary orbits. They are therefore inadequate in the case of binary stars (W. D. Heintz, 1971; R. G. Aitken, 1935).

Later, Sir John Herschel (1832) communicated a basically geometric method to the Royal Astronomical Society. Herschel's method was designed to utilize all the available data, so far as he considered them reliable. Since then, the contributions to the subject have been many. Some consist of entirely new methods of attack, others of modifications of those already proposed. Among the more notable investigators are Yvon Villarceau, H. H. Mädler, E. F. W. Klinkerfues, T. N. Thiele, M. Kowalsky, S. Glasenapp, H. J. Zwiers, K. Schwarzchild, T. J. J. See, H. N. Russell, R. T. A. Innes, W. H. van den Bos and others.

One may classify the various methods published so far into "geometric" methods, which are those that enforce only the constraint that the orbit is elliptical, and the "dynamical" ones which enforce, in addition, the area theorem.

The geometric treatment initiated by J. Herschel peaked in Zwiers' method (1896) and its modifications, e.g. those of H. M. Russell (1898) and of G. C. Comstock (1918). Every geometric method has the shortcoming that it must assume the location of the center of the ellipse to be known while it ignores the area theorem and thus fails to enforce one of the constraints to which the observations are subject. The growing quantity and quality of observations called for suitable computing precepts, and the successful return to dynamical methods began with van den Bos (1926).

Of the many methods for orbit-computation formulated, some are very useful and applicable to a wide range of problems, e.g. those by Zwiers, Russell and those by Innes and van den Bos.

Zwiers' method (1896) is essentially graphical and assumes that the apparent orbit has been drawn. This method is therefore useless unless the apparent ellipse gives a good geometrical representation of the observations and satisfies the law of areas, and thus will not be further described here since we are primarily concerned with the analytical methods.

In the following, we will briefly review Kowalsky's method and that by Thiele and Innes.

Definition of the Orbital Parameters

Seven parameters define the orbit and the space orientation of its plane. The first three of these (P,T,e) are dynamical and define the motion in the orbit; the last four (a,i,Ω,ω) are geometrical and give the size and orientation of the orbit. The parameters are defined somewhat differently from those for the orbits of planets and comets.

The first dynamical parameter P is the period of revolution, usually in units of mean sidereal years; n is the mean (usually annual) angular motion; since $n=2\pi/P$, P and n are equivalent. The second, T, is the epoch of periastron passage (usually expressed in terms of years and fractions thereof). The third, e, is the eccentricity of the orbital ellipse.

The geometrical parameter a is the angle subtended by the semi-major axis of the orbital ellipse (usually expressed in units of arcseconds). The angle i is the inclination of the orbital plane to the plane normal to the line of sight, that is, the angle between the plane of projection and that of the orbit in space. It ranges from 0° to 180°. When the position angle increases with time, that is, for direct motion, i is between 0° and 90°; for retrograde motion, i is counted between 90° and 180°;

and i is 90° when the orbit appears projected entirely onto the line of nodes. The "node", Ω, is the position angle of the line of intersection between the tangential plane of projection and the orbital plane. There are two nodes whose corresponding values of Ω differ by 180°. That node in which the orbital motion is directed away from the sun is called the ascending node. We understand Ω, which ranges from 0° to 360°, to refer to the ascending node. Because it is, however, one of the peculiarities of the orbit-determination of a visual binary that it is in principle impossible--from positional data alone--to decide whether the node is the ascending or descending one Ω may be restricted to $0° \leq \Omega \leq 180°$. The last, ω, is the longitude of the periastron in the plane of the orbit, counted positive in the direction of the orbital motion and starting at the ascending node. It ranges from 0° to 360°.

These definitions are adhered to throughout our work. Some of them may somewhat differ a little from those given by previous authors. But any way in which one defines them does not affect the principles of the method we describe.

The Method of M. Kowalsky and S. Glasenapp

This old method was first introduced by J. Herschel in a rather cumbersome form and is better known in its more direct formulation by M. Kowalsky in 1873 and by S. Glasenapp in 1889. R. G. Aitken (1935) gives the

detailed derivation of the formulae in his textbook The Binary Stars.

Kowalsky's method is essentially analytical. It derives the orbit parameters from the coefficients of the general equation of the ellipse which is the orthogonal projection of the orbit in space, the origin of coordinates being taken at the primary. The projected orbit can be expressed by a quadratic in x and y, whose five coefficients are related to those five orbital parameters which do not involve time.

In rectangular coordinates, the equation of an ellipse, and thus of a conic section, takes the form

$$c_1x^2 + 2c_2xy + c_3y^2 + 2c_4x + 2c_5y + 1 = 0 \quad , \qquad (2\text{-}1)$$

where the rectangular coordinates (x,y) are related to the more commonly directly observed polar coordinates (ρ,θ) by the equations

$$x = \rho\cos\theta \quad , \qquad (2\text{-}2a)$$

$$y = \rho\sin\theta \quad , \qquad (2\text{-}2b)$$

where ρ is the measured angular distance and θ the position angle.

The five coefficients of equation (2-1) can be determined by selecting five points on the ellipse or by all the available observations in the sense of least squares.

There is an unambiguous relationship between the five coefficients (c_1,c_2,c_3,c_4,c_5) and the five orbital parameters (a,e,i,ω,Ω). We can find the detailed derivation of the formulae in Aitken's The Binary Stars. Here, we will therefore state only the final formulae without derivation.

The five orbital parameters (a,e,i,ω,Ω) can be calculated from the known coefficients (c_1,c_2,c_3,c_4,c_5) by the following procedure.

1) The parameter Ω can be found from the equation

$$\tan 2\Omega = \frac{2(c_2-c_4c_5)}{(c_5^2-c_4^2+c_1-c_3)} \quad . \tag{2-3}$$

To determine in which quadrant Ω is located, we can use two other equations:

$$c'\sin 2\Omega = 2(c_2-c_4c_5) \quad , \tag{2-3'a}$$

$$c'\cos 2\Omega = c_5^2-c_4^2+c_1-c_3 \quad , \tag{2-3'b}$$

where

$$c' = \frac{\tan^2 i}{p^2} \quad , \tag{2-3'c}$$

which is always positive.

More elegantly, we write in Eichhorn's notation,

$$2\Omega = \text{plg}[2(c_2-c_4c_5),c_5^2-c_4^2+c_1-c_3] \quad . \qquad (2\text{-}3'\text{d})$$

H. Eichhorn (1985), in his "Kinematic Astronomy" (unpublished lecture notes), defines the plg(x,y) function as follows:

$$\text{plg}(x,y) = \arctan(x/y) + 90°[2-\text{sgn}x(1+\text{sgn}y)] \quad ,$$

where arctan is the principal value of the arctangent.

2) The inclination i is found from

$$\tan^2 i = -2 + \frac{2\cos\Omega(c_5^2+c_4^2-c_1-c_3)}{\cos 2\Omega(c_5^2+c_4^2-c_1-c_3)-(c_5^2-c_4^2+c_1-c_3)} \quad . \qquad (2\text{-}4)$$

Whether the i is in the first or second quadrant is determined by whether the motion is direct or retrograde. If the motion is direct, the position angle increases with time, $i<90°$; otherwise, $i>90°$.

3) The equation

$$\omega = \text{plg}[1-(c_4\sin\Omega-c_5\cos\Omega)\cos i, c_4\cos\Omega+c_5\sin\Omega) \qquad (2\text{-}5)$$

gives the value of ω.

4) With i,ω,Ω known, two more parameters (a and e) can be calculated from

$$e = \frac{2\cos 2\Omega(c_4 \sin\Omega - c_5 \cos\Omega)\cos i}{[\cos 2\Omega(c_5^2 + c_4^2 - c_1 - c_3) - (c_5^2 - c_4^2 + c_1 c_3)]\sin\omega} \quad , \tag{2-6}$$

$$a = \frac{2\cos 2\Omega}{[\cos 2\Omega(c_5^2 + c_4^2 c1 - c3) - (c_5^2 - c_4^2 + c1 - c3)](1 - e^2)} \quad . \tag{2-7}$$

5) To complete the solution analytically, the mean motion n (or the period P) and the time of periastron passage T, must be found from the mean anomaly M, computed from the observations by Kepler's equation:

$$M = n(t-T) = E - e\sin E \quad , \tag{2-8}$$

where E is the eccentric anomaly. Every M will give an equation of the form

$$M = nt + \Gamma \quad , \tag{2-9}$$

where $\Gamma = -nT$.

From these equations the values of n and T are computed by the method of least squares.

This is essentially the so-called Kowalsky method. It looks mathematically elegant. But because it is not only severely affected by uncertainties of observations but also ignores the area theorem, it has a very poor reputation among seasoned practitioners. However, in our work, we use

it for getting the initial approximation to the solution. It serves this purpose very well.

The Method of T. N. Thiele, R. T. A. Innes and W. H. van den Bos

T. N. Thiele (1883) published a method of orbit computation depending upon three observed positions and the constant of areal velocity.

The radii vectors to two positions in an orbit subtend an elliptical sector and a triangle, the sector being related to the time interval through the law of areas. Gauss introduced the use of the ratio: "sector to triangle" between different positions into the orbit computation of planets, and Thiele applied the idea to binary stars. Although the method could have been applied in a wide range of circumstances, it became widely used only after Innes and van den Bos revived it.

In 1926, R. T. A. Innes (Aitken, 1935), seeking a method simpler than those in common use for correcting the preliminary parameters of an orbit differentially, independently developed a method of orbit computation which differs from Thiele's in that he used rectangular instead of polar coordinates. W. H. van den Bos's (1926, 1932) merit is not merely a modification of the method (transcribing it for the use with Innes constants) but chiefly in its pioneeringly successful application. The device became most widely applied. Briefly, the method of computation is as follows.

The computation utilizes three positions (ρ_i,θ_i) or the corresponding rectangular coordinates (x_i,y_i) at the times t_i (i=1,2,3). The area constant C is the seventh quantity required. Thiele employs numerical integration to find the value of C.

First, from the observed data, find the quantities L by

$$L_{12} = t_2-t_1-D_{12}/C \quad , \tag{2-10a}$$

$$L_{23} = t_3-t_2-D_{23}/C \quad , \tag{2-10b}$$

$$L_{13} = t_3-t_1-D_{13}/C \quad , \tag{2-10c}$$

where $D_{ij} = \rho_i\rho_j\sin(\theta_j-\theta_i)$, are the areas of the corresponding triangles.

Then, from the equations

$$nL_{12} = p-\sin p \quad , \tag{2-11a}$$

$$nL_{23} = q-\sin q \quad , \tag{2-11b}$$

$$nL_{13} = (p+q)-\sin(p+q) \quad , \tag{2-11c}$$

the quantities n, p and q can be found by trials, for in the three equations above there are only three unknowns, i.e., n, p and q. The eccentric anomaly E_2 and the eccentricity e can thus be computed from

$$e\sin E_2 = \frac{(D_{23}\sin p-D_{12}\sin q)}{(D_{23}+D_{12}-D_{13})} \quad , \tag{2-12a}$$

$$e\cos E_2 = \frac{(D_{23}\cos p + D_{12}\cos q - D_{13})}{(D_{23} + D_{12} - D_{13})} \qquad (2\text{-}12b)$$

After E_2 and e are obtained, the two other eccentric anomalies E_1 and E_3 can be found from

$$E_1 = E_2 - p \quad , \qquad (2\text{-}13a)$$

$$E_3 = E_2 + q \quad . \qquad (2\text{-}13b)$$

The E_i are used first to compute the mean anomalies M_i from equations (2-8), which lead to three identical results for T as a check, and second to compute the coordinates X_i and Y_i from

$$X_i = \cos E_i - e \quad , \qquad (2\text{-}14a)$$

$$Y_i = \sin E_i \sqrt{(1-e^2)} \quad . \qquad (2\text{-}14b)$$

By writing

$$x_i = AX_i + FY_i \quad , \qquad (2\text{-}15a)$$

$$y_i = BX_i + GY_i \quad , \qquad (2\text{-}15b)$$

the constants A, B, F, G are obtained from two positions, the third again serving as a check. These four coefficients, A, B, F, G, are the now so-called Thiele-Innes constants. In addition to n, T, e, which have already been

determined, the other four orbital parameters a, i, Ω, ω can be calculated from A, B, F, G through

$$\Omega+\omega = \mathrm{plg}(B-F,A+G) \quad , \tag{2-16a}$$

$$\Omega-\omega = \mathrm{plg}(B+F,A-G) \quad , \tag{2-16b}$$

$$\tan^2\frac{i}{2} = \frac{(A-G)\cos(\omega+\Omega)}{(A+G)\cos(\omega-\Omega)} = -\frac{(B+F)\sin(\omega+\Omega)}{(B-F)\sin(\omega-\Omega)} \quad , \tag{2-16c}$$

$$A+G = 2a\cos(\omega+\Omega)\cos^2\frac{i}{2} \quad . \tag{2-16d}$$

In addition to the brief introduction to this method given above, a detailed description of it can be found in many books, e.g., Aitken's book The Binary Stars (1935) and W. D. Heintz's book Double Stars (1971).

A more accurate solution is obtained by correcting differentially the preliminary orbit which was somehow obtained by using whatever method. This correction can be achieved by a least squares solution.

The Method of Least Squares

The method of least squares was invented by Gauss and first used by him to calculate orbits of solar system bodies from overdetermined system of equations. It is the most important tool for the reduction and adjustment of observations in all fields, not only in astronomy. However, the traditional standard algorithm, which employs condition equations that are solved with respect to the (uncorrelated)

observations and either linear in the adjustment parameters or linearized by developing them in Taylor series which are broken off after the first order terms, is inadequate for treating the problem at hand. An algorithm for finding the solution of a more general least squares adjustment problem was given by D. C. Brown (1955). This situation may briefly be described as follows.

Let {x} be a set of quantities for which an approximation set $\{x_0\}$ was obtained by direct observation. By ordering the elements of the sets {x} and $\{x_o\}$ and regarding them as vectors, $\mathbf{x}$ and $\mathbf{x_0}$, respectively, the vector $\mathbf{v}=\mathbf{x}-\mathbf{x_0}$ is the vector of the observational errors which are initially unknown. Assume that they follow a multivariate normal distribution and that their covariance matrix σ is regarded as known. Further assume that a set of parameters {a} is ordered to form the vector a. The solution of the least squares problem (or the adjustment) consists in finding those values of the elements of {x} and {a} which minimize the quadratic form $\mathbf{v}^T\sigma^{-1}\mathbf{v}$ while at the same time rigorously satisfying the condition equations

$$f_i(\{x\}_i,\{a\}_i)=0 \quad . \tag{2-17}$$

This is a problem of finding a minimum of the function $\mathbf{v}^T\sigma^{-1}\mathbf{v}$ subject to condition equations. A general rigorous and noniterative algorithm for the solution exists only for

the case that the elements of {x} and {a} occur linearly in the functions f_i. When f_i are nonlinear in the elements of either {x} or {a}, or both, equations which are practically equivalent to the f_i and which are linear in the pertinent variables can be derived in the following way.

Define a vector δ by $\mathbf{a}=\mathbf{a}_0+\delta$. The condition equations can then be written

$$\mathbf{f}(\mathbf{x}_0+\mathbf{v},\ \mathbf{a}_0+\delta) = \mathbf{0} \quad , \tag{2-18}$$

where the vector of functions $\mathbf{f}=(f_1, f_2 \ldots, f_m)^T$, m being the number of equations for which observations are available. Now assume that all elements of {v} and {δ} are sufficiently small so that the condition equations can be developed as a Taylor series and written

$$\mathbf{f}_0 + \mathbf{f}_x\mathbf{v} + \mathbf{f}_a\delta + O(2) \ldots = \mathbf{0} \quad , \tag{2-18'a}$$

where $\mathbf{f}_0 = \mathbf{f}(\mathbf{x}_0,\mathbf{a}_0)$, $\mathbf{f}_x = \left(\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right)\Big|_{\mathbf{x}_0,\mathbf{a}_0}$ and $\mathbf{f}_a = \left(\frac{\partial \mathbf{f}}{\partial \mathbf{a}}\right)\Big|_{\mathbf{x}_0,\mathbf{a}_0}$.

If the small quantities of order higher than 1 can be neglected, we can write the linearized condition equations as

$$\mathbf{f}_0 + \mathbf{f}_x\mathbf{v} + \mathbf{f}_a\delta = \mathbf{0} \quad . \tag{2-18'b}$$

These are linear in the relevant variables, which are the components of $\mathbf{v}$ and of δ.

In order to satisfy the conditions (2-18'b), we define a vector -2μ of Lagrangian multipliers and minimize the scalar function

$$S(\mathbf{v},\delta) = \mathbf{v}^T\sigma^{-1}\mathbf{v} - 2\mu(\mathbf{f}_0 + \mathbf{f}_x\mathbf{v} + \mathbf{f}_a\delta) \qquad (2\text{-}19)$$

in which the components of $\mathbf{v}$ and δ are the variables. Setting the derivatives $(\partial S/\partial v)$ and $(\partial S/\partial \delta)$ equal to zero and considering equations (2-18'b), we obtain

$$\delta = - [\mathbf{f}_a^T(\mathbf{f}_x\sigma\mathbf{f}_x^T)\mathbf{f}_a]^{-1}\mathbf{f}_a^T(\mathbf{f}_x\sigma\mathbf{f}_x^T)^{-1}\mathbf{f}_0 \quad , \qquad (2\text{-}20a)$$

$$\mu = - (\mathbf{f}_x\sigma\mathbf{f}_x^T)^{-1}(\mathbf{f}_a\delta+\mathbf{f}_0) \quad , \qquad (2\text{-}20b)$$

$$\mathbf{v} = \sigma\mathbf{f}_x^T\mu \quad , \qquad (2\text{-}20c)$$

where we have assumed that $\mathbf{f}_x\sigma\mathbf{f}_x^T$ is nonsingular. This is the case only if all equations $f_i=0$ contain at least one component of $\mathbf{x}$; i.e., if there are no pure constraints of the form $g_i(a)=0$. This case (which we shall not encounter in our investigations) is discussed below in the description of Jefferys' method.

Note that in constructing the scalar function S in expression (2-19), first order approximations have

been used. In some cases, the linearized representation of Eq. (2-18) by Eq. (2-18'b) is not accurate enough. In some of these cases, the convergence toward the definitive solution may be accelerated and sometimes be brought about by a method suggested by Eichhorn and Clary (1974) when a strictly linear approach would be divergent. Their solution algorithm takes into account the second (as well as the first) order derivatives in the adjustment residuals (observational errors) and the corrections to the initially available approximations to the adjustment parameters. They pointed out that the inclusion of second order terms in the adjustment residuals is necessary whenever the adjustment residuals themselves cannot be regarded as negligible as compared to the adjustment parameters, in which cases the conventional solution techniques would not lead to the "best" approximations for the adjustment parameters in the sense of least squares. The authors modified the condition equations as

$$\mathbf{f}_0 + \mathbf{f}_x\mathbf{v} + \mathbf{f}_a\boldsymbol{\delta} + \mathbf{V}\mathbf{v} + \mathbf{D}\boldsymbol{\delta} + \begin{matrix}\mathbf{M}\boldsymbol{\delta}\\ \text{or}\\ \mathbf{N}\mathbf{v}\end{matrix} = \mathbf{0} \quad . \qquad (2\text{-}18")$$

Correspondingly, the scalar function to be minimized becomes

$$S"(\mathbf{v},\boldsymbol{\delta}) = \mathbf{v}^T\boldsymbol{\sigma}^{-1}\mathbf{v} - 2\boldsymbol{\mu}(\mathbf{f}_0+\mathbf{f}_x\mathbf{v}+\mathbf{f}_a\boldsymbol{\delta}+\mathbf{V}\mathbf{v}+\mathbf{D}\boldsymbol{\delta}+\begin{matrix}\mathbf{M}\boldsymbol{\delta}\\ \text{or}\\ \mathbf{N}\mathbf{v}\end{matrix}) \quad . \quad (2\text{-}19")$$

Here, the i-th line of the matrix **D** is $\frac{1}{2}\sigma^T \mathbf{E}_i$, and

$$E_i = \left(\frac{\partial^2 f_i}{\partial a_j \partial a_k} \right) \Bigg|_{x_0, a_0} \quad ; \tag{2-21}$$

the matrix of the Hessean determinant of f_i with respect to the adjustment parameters. Similarly, the i-th line of vector **V** is $\frac{1}{2}\mathbf{v}^T \mathbf{W}_i$, and

$$W_i = \left(\frac{\partial^2 f_i}{\partial x_j \partial x_k} \right) \Bigg|_{x_0, a_0} \quad ; \tag{2-22}$$

the i-th line of **M** is $\mathbf{v}^T \mathbf{H}_i$, and

$$H_i = \left(\frac{\partial^2 f_i}{\partial x_j \partial a_k} \right) \Bigg|_{x_0, a_0} \quad ; \tag{2-23}$$

and that of **N** is $\delta^T \mathbf{H}_i^T$, so evidently $\mathbf{M}\delta = \mathbf{N}\mathbf{v}$.
Minimizing S", also δ, **v** can be calculated.

For this algorithm in detail, one can refer to the original papers of Eichhorn and Clary. The notation used here is slightly different from the original one used by the authors.

Jefferys (1980, 1981) proposed an even more accurate algorithm which also improves the convergence of the

conventional least squares method. Furthermore, his method is nearly as simple to apply in practice as the classical method of least squares, because it does not require any second order derivatives. Jefferys defines the scalar function to be minimized as

$$S = \frac{1}{2}\mathbf{v}^T\sigma^{-1}\mathbf{v} + \mathbf{f}^T(\hat{\mathbf{x}},\hat{\mathbf{a}})\hat{\mu} + \mathbf{g}^T(\hat{\mathbf{a}})\,\hat{\gamma} \quad , \tag{2-24}$$

where $\hat{\mathbf{x}}=\mathbf{x}_0+\hat{\mathbf{v}}$; $\hat{\mathbf{x}},\hat{\mathbf{a}}$ are the current "best" approximations to $\mathbf{x}$ and $\mathbf{a}$; and $\mathbf{g}$ is another vector function consisting of the constraints on the parameters; $\hat{\mu}$, $\hat{\gamma}$ are vectors of Langrangian multipliers and evaluated at $(\hat{\mathbf{x}}, \hat{\mathbf{a}})$. Minimizing S with respect to $\hat{\mathbf{x}}$ and $\hat{\mathbf{a}}$, he arrives at the normal equations

$$\sigma^{-1}\hat{\mathbf{v}} + \mathbf{f}_{\hat{\mathbf{x}}}^T(\hat{\mathbf{x}},\hat{\mathbf{a}})\hat{\mu} = 0 \quad , \tag{2-25a}$$

$$\mathbf{f}_{\hat{\mathbf{x}}}^T(\hat{\mathbf{x}},\hat{\mathbf{a}})\hat{\mu} + \mathbf{g}_{\hat{\mathbf{a}}}^T(\hat{\mathbf{a}})\hat{\gamma} = 0 \quad , \tag{2-25b}$$

$$\mathbf{f}(\hat{\mathbf{x}},\hat{\mathbf{a}}) = 0 \quad , \tag{2-25c}$$

$$\mathbf{g}(\hat{\mathbf{a}}) = 0 \quad , \tag{2-25d}$$

where

$$\mathbf{f}_{\hat{\mathbf{x}}} = \mathbf{f}_{\mathbf{x}}\Big|_{\hat{\mathbf{x}},\hat{\mathbf{a}},} \qquad \mathbf{f}_{\hat{\mathbf{a}}} = \mathbf{f}_{\mathbf{a}}\Big|_{\hat{\mathbf{x}},\hat{\mathbf{a}}.}$$

These equations are exact and involve no approximations. This is the significant difference between Jefferys' method and those of Brown and of Eichhorn and Clary.

Remarks

As mentioned before, Kowalsky's method will most likely not produce the best obtainable results, because the relative observed coordinates (x, y, t) are subjected only to the condition (2-1), which involves only five of the seven necessary orbital parameters as adjustment parameters and does not enforce the area theorem. It can therefore never be used for a definitive orbit determination since it completely ignores the observation epochs.

Yet, Eq. (2-1) has the advantage that it appears to be simple, in particular it is linear in the adjustment parameters albeit not in the observations. When it is used for the determination of orbits, the right-hand sides of the equations which result from inserting a pair (x, y) of observed rectangular coordinates into Eq. (2-1) are regarded as errors with a univariate Gaussian distribution (i.e., as normally distributed errors). One may then perform a least squares adjustment which is linear in the adjustment parameters. As Eichhorn (1985) pointed out, this approach, while it has the advantage that approximation values for the adjustment parameters need not be available at the outset, fails to take into account two facts.

1) It is not the right-hand sides of the condition equations which are to be considered as normally distributed errors, but rather the observations (x,y) or (ρ,θ). The condition equations (2-1) thus contain more than one observation each.

Since the observations occur in the condition equations nonlinearly, the matrix

$$\mathbf{f_x} = \left(\frac{\partial \mathbf{f(x,a)}}{\partial \mathbf{x}} \right) \Bigg|_{\mathbf{x}=\mathbf{x}_0, \mathbf{a}=\mathbf{a}_0}$$

must be found. This requires knowledge of approximate values $\mathbf{a_0}$ for **a**. Approximate values $\mathbf{x_0}$ for **x** are available--they are the observations themselves. Approximate values $\mathbf{a_0}$ for **a** may sometimes indeed be obtained in the classical way by regarding the right-hand sides of the condition equations as normally distributed errors. In addition, it should also be taken into account that the covariance matrix σ of the observations is not necessarily diagonal.

2) In some cases, especially when the binary under study is very narrow, the errors of the observations are not negligibly small compared with the adjustment parameters. This requires either that second-order terms in the observational errors **v** be carried in the equations or, as Jefferys has pointed out, that iterations be performed using in the evaluation of the matrices $\mathbf{f_x}$ and $\mathbf{f_a}$ not only improved approximations for **a** but also improved values for the observed quantities as they become available.

If Kowalsky's methods were so modified, the algorithm would yield better values for the adjustment parameters **a**

than the traditional approach. Either way, one can usually find an initial approximation by Kowalsky's method.

With respect to both the theoretical clarity and the practical applicability, as far as it is concerned, the Thiele-Innes-van den Bos method leaves nothing to be desired. However, the three places selected, even when smoothed graphically or by some computation, may not suffice to describe the motion with sufficient accuracy, so that large and systematic residuals may remain, particularly near periastron. The method is seriously inadequate even if one of the ratios sector to triangle is very close to 1 and thus strongly affected by the measurement errors or if the area constant C is not initially known to the required accuracy. The computation may then produce an erroneous orbit with spuriously high eccentricity, perhaps a hyperbolic one, or no solution at all. And obviously, different combinations of the three positions selected from a set of observations will not likely give the same result. This method therefore fails to use the information contained in the observations in the best possible way.

In our work we try to present a fairly general least squares algorithm to solve the orbit problem. We shall adopt Jefferys' least squares method as our basic approach.

CHAPTER III
GENERAL CONSIDERATIONS

This chapter contains a general discussion of the least squares orbit problem. We shall set up condition equations which simultaneously satisfy the ellipse equation and the area theorem.

We have seen that it is not sufficient to use Eq. (2-1) as the only type of condition equation because this would ignore the observing epochs, cf. last chapter. Completely appropriate condition equations must explicitly contain the complete set of the seven independent orbital parameters as the adjustment parameters. Also, to be useful in practice, they must impose both the geometric and dynamical conditions, and must lead to a convergent sequence of iterations.

After the condition equations are established, we present the general outline of the algorithm which solves the orbit problem by Jefferys' method of least squares.

We also discuss some further suggestions for obtaining the initial approximate solution required for the least squares algorithm.

The Condition Equations

Assume that a set of observations $\{x_0\}$ was obtained consisting of complete data triples (t, ρ, θ), which measure the positions of the fainter component (secondary) with respect to the brighter one (primary): the position angle θ is counted counterclockwise from North and ranges from 0° to 360°; the angular separation ρ (also called distance) is usually given in seconds of arc, and t is the observing epoch. The conversion of (ρ, θ) to rectangular coordinates (x, y) in seconds of arc is as following:

Declination difference $\delta_c - \delta_p = x = \rho\cos\theta$, (3-1a)

Right ascension difference $(\alpha_c - \alpha_p)\cos\delta = y = \rho\sin\theta$, (3-1b)

where δ_c, α_c are the declination and right ascension, respectively, of the secondary; δ_p, α_p those of primary.

Equivalently, the observations can also be regarded as relative coordinates (t, x, y) of the secondary with respect to the primary.

It might be worthwhile to point out that 1) even though the formulae (3-1) are approximations valid only for small values of ρ, they may be regarded as practically rigorous for binaries; 2) we are following the custom in double star

astronomy by having the x-axis along the colure* and the y-axis tangential to the parallel of declination.

All observations in $\{x_0\}$ are affected by observational errors. Let $\{x\}$ be the set of the true values of the observations, that is, those values the observations would have had if there were no observational errors. By ordering the elements of the sets $\{x_0\}$ and $\{x\}$ and regarding them as vectors, $\mathbf{x}_0$ and $\mathbf{x}$ respectively, we have seen that we may write the vector of observational errors as $\mathbf{v} = \mathbf{x} - \mathbf{x}_0$. These errors are of course unknown, but as mentioned already in the last chapter, we assume that they follow a multivariate normal distribution with known covariance matrix. For visual binaries, the relative orbit must be an ellipse (strictly speaking, a conic section) in space as well as in projection. All pairs (x, y) must therefore satisfy the condition equations (2-1):

$$C_1x^2 + 2C_2xy + C_3y^2 + 2C_4x + 2C_5y + 1 = 0 \quad ,$$

which implicitly involve five of the seven orbital parameters but do not enforce the area theorem.

The well-known relationships between the five coefficients (C_1, C_2, C_3, C_4, C_5) in Eq. (2-1) and the five

---

*Following Eichhorn's terminology who uses the term "colure" generally for any locus of constant right ascension.

orbital parameters (e, a, i, ω, Ω) by way of the Thiele-Innes constants, have been discussed in the last chapter.

Consider a right-handed astrocentric coordinate system **K** whose X-Y plane is the true orbital plane such that the positive X-axis points toward the periastron (of the secondary with respect to the primary). The positive Y-axis is obtained by rotating the X-axis by 90° on the Z-axis in the direction of the orbital motion. The axes of a second astrocentric, right-handed coordinate system **k** are parallel to those of the equator system **Q**. The two systems **K** and **k** are related by the transformation

$$\mathbf{x}^{\mathbf{K}} = \mathbf{R}_3(\omega)\mathbf{R}_1(i)\mathbf{R}_3(\Omega)\mathbf{X}^{\mathbf{k}} \quad . \tag{3-2}$$

From the theory of the two-body problem we know that, in the system **K**, the coordinates of the secondary with respect to the primary are given by

$$\mathbf{x}^{\mathbf{K}} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = a\begin{pmatrix} \cos E - e \\ \sin E\sqrt{1-e^2} \\ 0 \end{pmatrix} \quad , \tag{3-3}$$

where E is the eccentric anomaly, which is the solution of Kepler's equation

$$n(t-T) = E - e\sin E \quad . \tag{3-4}$$

Here, n and T, the mean motion and the periastron epoch, are the two orbital parameters not involved in Eq. (2-1).

From Eq. (3-2) we obtain

$$x^K = \begin{pmatrix} A & B & C \\ F & G & H \\ K & L & M \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} , \tag{3-5}$$

where

$$z = - \frac{Kx + Ly}{M} , \tag{3-6}$$

and

$$\begin{pmatrix} A & B & C \\ F & G & H \\ K & L & M \end{pmatrix} = R_3(\omega)R_1(i)R_3(\Omega) , \tag{3-7}$$

or, in detail,

$$A = \cos\Omega\cos\omega - \sin\Omega\sin\omega\cos i \quad ; \tag{3-8a}$$

$$B = \sin\Omega\cos\omega + \cos\Omega\sin\omega\cos i \quad ; \tag{3-8b}$$

$$C = \sin\omega\sin i \quad ; \tag{3-8c}$$

$$F = -\cos\Omega\sin\omega - \sin\Omega\cos\omega\cos i \quad ; \tag{3-8d}$$

$$G = -\sin\Omega\sin\omega + \cos\Omega\cos\omega\cos i \quad ; \tag{3-8e}$$

$$H = \cos\omega\sin i \quad ; \tag{3-8f}$$

$$K = \sin\Omega\sin i \quad ; \tag{3-8g}$$

$$L = -\cos\Omega\sin i \quad ; \tag{3-8h}$$

$$M = \cos i \quad , \tag{3-8i}$$

where in this notation, the traditional Thiele-Innes constants would be aA, aB, aF and aG.

From Eq. (3-7) and (3-5) we can get

$$X = Ax + By + Cz = \frac{Gx - Fy}{M} \quad ; \tag{3-9a}$$

$$Y = Fx + Gy + Hz = -\frac{Bx - Ay}{M} \quad ; \tag{3-9b}$$

Thus, we see that X and Y can be expressed in terms of i,ω,Ω and the observations (x, y).

From Eq. (3-3) we obtain

$$\cos E = \frac{X}{a} + e \quad , \tag{3-10a}$$

$$\sin E = \frac{Y}{a\sqrt{1-e^2}} \quad , \tag{3-10a}$$

Combining (3-10) with Kepler's equation (3-4), we get

$$\frac{X}{a} + e = \cos\left[n(t-T) + \frac{eY}{a\sqrt{1-e^2}}\right] \quad , \tag{3-11a}$$

$$\frac{Y}{a\sqrt{1-e^2}} = \sin\left[n(t-T) + \frac{eY}{a\sqrt{1-e^2}}\right] \quad , \qquad (3\text{-}11b)$$

More succinctly, we have

$$\mathbf{U} = \cos[\, n(t-T) + e\mathbf{V} \,] \quad , \qquad (3\text{-}12a)$$

$$\mathbf{V} = \sin[\, n(t-T) + e\mathbf{V} \,] \quad , \qquad (3\text{-}12b)$$

or

$$\mathbf{U} = \cos[\, n(t-T) + e\sqrt{1-\mathbf{U}^2} \,] \quad , \qquad (3\text{-}12'a)$$

$$\mathbf{V} = \sin[\, n(t-T) + e\mathbf{V} \,] \quad , \qquad (3\text{-}12'b)$$

with

$$\mathbf{U} = \frac{X}{a} + e \quad , \qquad \mathbf{V} = \frac{Y}{a\sqrt{1-e^2}} \quad . \qquad (3\text{-}12c)$$

After X and Y are expressed in terms of i, ω, Ω and (x, y) as in Eqs. (3-9), Eqs. (3-11) or (3-12) involve exactly the seven orbital parameters (n, T, a, e, i, ω, Ω) and the observations (t, x, y). The observing epoch t now appears explicitly, as it must if the area theorem is to be enforced.

Now, we see that if both equations (3-11) are satisfied, Kepler's equation which enforces area theorem would be satisfied and furthermore, the ellipse equation would also be automatically satisfied as can be seen if t is eliminated from Eqs. (3-11) so that these equations are reduced to one equation in X and Y. If we select the two equations (3-11)

as the condition equations, we need no longer carry Eq. (2-1) separately. Of course, we can use any one of Eqs. (3-11) as well as Eq. (2-1) as the condition equations. However, using Eq. (2-1) is not convenient, for it contains the five coefficients directly, but not the five orbital parameters themselves, even though there are unique relationships between them. Ideally, the condition equations should have the adjustment parameters explicitly as variables.

In our work, we use the Eqs. (3-11) as the complete set of condition equations.

The General Statement of the Least Squares Orbit Problem

As seen in the last section, the vector of "true observations" $\mathbf{x}$ (presumably having been adjusted from the observation $\mathbf{x}_0$ by the residuals $\mathbf{v}$) and the vector of "true orbital parameters" a, $a=(n, T, a, e, i, \omega, \Omega)^T$, must satisfy the condition equations

$$f_1(\mathbf{x}_0+\mathbf{v},\mathbf{a}) = \frac{X}{a} + e - \cos\left[n(t-T)+ \frac{eY}{a\sqrt{1-e^2}}\right] = 0 \quad , \qquad (3\text{-}13a)$$

$$f_2(\mathbf{x}_0+\mathbf{v},\mathbf{a}) = \frac{Y}{a\sqrt{1-e^2}} - \sin\left[n(t-T)+ \frac{eY}{a\sqrt{1-e^2}}\right] = 0 \qquad (3\text{-}13b)$$

where

$$\begin{pmatrix} X \\ Y \\ 0 \end{pmatrix} = \begin{pmatrix} A & B & C \\ F & G & H \\ K & L & M \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} ,$$

with

$$z = - \frac{Kx + Ly}{M} ,$$

and

$$\begin{pmatrix} A & B & C \\ F & G & H \\ K & L & M \end{pmatrix} = \mathbf{R}_3(\omega)\mathbf{R}_1(i)\mathbf{R}_3(\Omega) \quad .$$

In our problem, there are no constraints between the parameters which involve no observations so that Jefferys' **g** function does not occur. The problem can therefore be stated as follows.

Assume that the residuals {v} (regarded as vector **v**) of a set of observations $\{x_0\}$ (regarded as vector $\mathbf{x}_0$) follow a multivariate normal distribution, whose covariance matrix $\boldsymbol{\sigma}$ is regarded as known; we are to find the best approximations of $\hat{\mathbf{v}}$ (for **v**, the residuals) and $\hat{\mathbf{a}}$ (for **a**, the parameters) such that

$$\mathbf{f}_1(\mathbf{x}_0+\hat{\mathbf{v}},\hat{\mathbf{a}}) = 0$$

and

$$\mathbf{f}_2(\mathbf{x}_0+\hat{\mathbf{v}},\hat{\mathbf{a}}) = 0$$

are both satisfied while at the same time the quadratic form

$$S_0 = \frac{1}{2} \hat{v}^T \sigma^{-1} \hat{v} \qquad (3\text{-}14)$$

is minimized.

Following the well-known procedure introduced by Lagrange, the solution is obtained by minimizing the scalar function

$$S = \frac{1}{2} \hat{v}^T \sigma^{-1} \hat{v} + f^T(\hat{x},\hat{a})\hat{\mu} \quad , \qquad (3\text{-}15)$$

where $\hat{x} = \hat{x}_0 + \hat{v}$, and $\hat{\mu}$ is the vector of Lagrangian multipliers, together with satisfying the equations $f_1=0=f_2$. Denoting the matrix of partial derivatives with respect to a variable by a subscript, this is equivalent to solving the following normal equations:

$$\sigma^{-1}\hat{v} + f_{\hat{x}}^T(\hat{x},\hat{a})\hat{\mu} = 0 \qquad (3\text{-}16a)$$

$$f_{\hat{a}}^T\hat{\mu} = 0 \quad , \qquad (3\text{-}16b)$$

$$f(\hat{x},\hat{a}) = 0 \quad . \qquad (3\text{-}16c)$$

We have stated before that these equations are exact and therefore involve no approximations. Before Jefferys, all authors used first order or second order approximations in forming S in equation (3-15). This is the significant

difference that distinguishes Jefferys' method from those employed by previous authors.

It is evident that the solution of the equations (3-16) would solve the posed problem.

About the Initial Solution

The least squares algorithm requires an initial solution as starting point. Any approach which leads to approximate values of the orbital parameters serves this purpose, because our algorithm does not require a very accurate initial approximation. As long as the initial approximation is not too different from the final result, convergence can always be achieved. To obtain an initial solution, the following procedures may lead to an initial solution.

1) As mentioned in Chapter II, Kowalsky's method can produce a preliminary solution. Inserting the pairs (x, y) of observed rectangular coordinates into Eq. (2-1), we have a set of linear equations in which the five coefficients (c_1, c_2, c_3, c_4, c_5) are the unknowns. By making a classical least squares solution based on these linear equations, the five coefficients can be computed. An estimate of the five parameters (a, e, i, ω, Ω) can be obtained in turn from the unique relationships between them and the five coefficients. The remaining two parameters (n, T) also can then be calculated from the known quantities simply by classical least squares, as described in Chapter II.

As Eichhorn (1985) has pointed out, it is of course better to use the modified Kowalsky method. Using (2-1) as condition equations and the five coefficients as the adjustment parameters, one may iterate by Jefferys' algorithm toward the best fitting adjustment parameters and the best corrections to the observations, $\mathbf{v}$, that is, to arrive at the values of $\hat{\mathbf{a}}$ and $\hat{\mathbf{v}}$ which minimize the scalar function S_0 (see equation 3-14) while simultaneously satisfying the condition equations.

This method is simple to apply in practice. But unfortunately, especially when the observations are not very precise, the five coefficients (c_1, c_2, c_3, c_4, c_5) in some cases do not always satisfy the conditions for an ellipse; i.e., they do not meet the requirements

$$c_1>0,\ c_3>0 \quad \text{and} \quad c_1c_3-c_2^2>0.$$

However, in these cases, it does not mean that there is no elliptic solution at all and other approaches can be tried. When this happens, one may for instance take the approach outlined in Chapter 23 of Lawson and Harsson (1974).

2) By using some selected points among the observation data instead of using all the data points, sometimes a solution can be found by Kowalsky's method. Such a solution is usually also good enough to be a starting approximation.

3) Or, carefully selecting three points among the observation data, one may use the Thiele-Innes-van den Bos method to calculate an initial approximation. The method has been described in Chapter II.

CHAPTER IV
THE SOLUTION BY NEWTON'S METHOD

The Solution by Newton's Method

In our problem, the normal equations (3-16) are nonlinear. They must be solved by linearization and successive approximations. Assume that approximate initial estimates of the unknowns in the normal equations have somehow been obtained (using whatever methods). This initial approximation may be improved by Newton's method, which consists of linearizing the normal equations about the available solution by a first order development and obtaining an improved solution by solving the linearized equations. This process is iterated with the hope that convergence to a definite solution would eventually be obtained. This expectation is reasonable if the initial approximation is sufficiently close to the final solution and if the observational errors are not too large.

Following Jefferys' notation (which is also the notation we have used in Chapter III), let the initial approximate solution (and also the current approximate solution during iteration) be given by $(\hat{\mathbf{x}},\hat{\mathbf{a}})$, where $\hat{\mathbf{x}} = \mathbf{x}_0 + \hat{\mathbf{v}}$, $\mathbf{x}_0$ being the vector of observation, $\hat{\mathbf{v}}$ the vector of observational errors (for which we adopt the nullvector

as initial approximation); $\hat{\mathbf{a}}$ is the initial approximation of the vector of the seven orbital parameters (adjustment parameters); also let the corrections to both $\hat{\mathbf{x}}$ and $\hat{\mathbf{a}}$ be denoted by $\hat{\boldsymbol{\epsilon}}$ and $\boldsymbol{\delta}$, respectively.

The normal equations in our problem now become

$$\sigma^{-1}(\hat{\mathbf{v}}+\hat{\boldsymbol{\epsilon}}) + \mathbf{f}_{\hat{\mathbf{x}}}^{\mathrm{T}}\hat{\boldsymbol{\mu}} = \mathbf{0} \quad , \tag{4-1a}$$

$$\mathbf{f}_{\hat{\mathbf{x}}}^{\mathrm{T}}\hat{\boldsymbol{\mu}} = \mathbf{0} \quad , \tag{4-1b}$$

$$\hat{\mathbf{f}} + \mathbf{f}_{\hat{\mathbf{x}}}\hat{\boldsymbol{\epsilon}} + \mathbf{f}_{\hat{\mathbf{a}}}\boldsymbol{\delta} = \mathbf{0} \quad . \tag{4-1c}$$

Here we have ignored (as Jefferys did) products of $\hat{\boldsymbol{\epsilon}}$ and $\boldsymbol{\delta}$ with Lagrangian Multipliers. This does not affect the final result, as Jefferys also pointed out. A caret in equations (4-1) above means evaluation at current values of $\hat{\mathbf{x}}$ and $\hat{\mathbf{a}}$.

Similar to Jefferys' procedure, we solve the equations (4-1) as follows.

Solving Eq. (4-1a) for $\hat{\boldsymbol{\epsilon}}$ we have

$$\hat{\boldsymbol{\epsilon}} = - \hat{\mathbf{v}} - \sigma\mathbf{f}_{\hat{\mathbf{x}}}\hat{\boldsymbol{\mu}} \quad . \tag{4-2}$$

Substituting (4-2) into (4-1c) for $\hat{\boldsymbol{\epsilon}}$, Eq. (4-1c) becomes

$$\hat{\mathbf{f}} - \mathbf{f}_{\hat{\mathbf{x}}}\hat{\mathbf{v}} - \mathbf{f}_{\hat{\mathbf{x}}}\sigma\mathbf{f}_{\hat{\mathbf{x}}}^{\mathrm{T}}\hat{\boldsymbol{\mu}} + \mathbf{f}_{\hat{\mathbf{a}}}\boldsymbol{\delta} = \mathbf{0} \quad . \tag{4-3}$$

Solving Eq. (4-3) for $\hat{\boldsymbol{\mu}}$, we obtain

$$\hat{\mu} = w(\hat{f} - f_{\hat{x}}\hat{v} + f_{\hat{a}}\delta) \quad , \tag{4-4}$$

where the "weight matrix" w is given by

$$w = (f_{\hat{x}}\sigma f_{\hat{x}}^T)^{-1} \quad . \tag{4-5}$$

Inserting this solution for $\hat{\mu}$ into Eq. (4-1b), we have

$$f_{\hat{a}}^T w(\hat{f} - f_{\hat{x}}\hat{v} + f_{\hat{a}}\delta) = 0 \quad . \tag{4-6}$$

If we now define

$$\hat{\phi} = \hat{f} - f_{\hat{x}}\hat{v} \quad , \tag{4-7}$$

and rearrange Eq. (4-6), the equation for δ will have the form

$$(f_{\hat{a}}^T w f_{\hat{a}})\delta = - f_{\hat{a}}^T w\hat{\phi} \quad . \tag{4-8}$$

This set of linear equations is easy to solve for the corrections δ by general methods. With this solution for δ, the improved residuals $\hat{v}_n$ are obtained from the equation

$$\hat{v}_n = - \sigma f_{\hat{x}}^T w(\hat{\phi}+f_{\hat{a}}\delta) \quad , \tag{4-9}$$

which follows from Eqs. (4-1a), (4-4) and (4-7). We then get the new vectors of $\hat{\mathbf{a}}_n$ and $\hat{\mathbf{x}}_n$ as

$$\hat{\mathbf{a}}_n = \hat{\mathbf{a}} + \boldsymbol{\delta} \quad , \tag{4-10a}$$

$$\hat{\mathbf{x}}_n = \mathbf{x}_0 + \hat{\mathbf{v}}_n \quad , \tag{4-10b}$$

which constitute the improved solution.

After each iteration, we check the relative magnitude of each component in $\hat{\mathbf{v}}$ and $\boldsymbol{\delta}$ against the corresponding component in $\hat{\mathbf{x}}$ and $\hat{\mathbf{a}}$ and get the maximum value among all of these and test if this value is smaller than some specified number, say 10^{-8}. If the improved solution is still not sufficiently accurate (i.e. if the above found maximum value is still not smaller than the specified value), the process of iterations has to be continued until convergence has been attained, that is, until subsequent iterations no longer give significant corrections.

At the outset, the obvious starting point for this scheme is to adopt $\hat{\mathbf{x}}=\mathbf{x}_0$ as the initial approximation for the "true observation" vector $\mathbf{x}$ (in this case, $\hat{\mathbf{v}}=\mathbf{0}$), and to use as first approximation of $\hat{\mathbf{a}}$ for $\mathbf{a}$ a vector $\hat{\mathbf{a}}_0$, an initial solution of the seven orbital parameters, which has been obtained somehow.

It is important for convergence that the initial solution of $(\hat{\mathbf{x}},\hat{\mathbf{a}})$ is not too different from the final solution which is obtained by the process given by Eq. (4-1)

through (4-10). In Chapter III, we discussed how to find a good approximation as an initial solution.

According to the scheme outlined above, the application of Newton's method in our problem would consist of the following steps.

Step 1.

Calculate $\mathbf{f}$, $\mathbf{f}_{\hat{\mathbf{x}}}$ and $\mathbf{f}_{\hat{\mathbf{a}}}$ from the current values of $\hat{\mathbf{x}}$ and $\hat{\mathbf{a}}$;

Step 2.

Calculate $\hat{\boldsymbol{\phi}}$ from $\hat{\boldsymbol{\phi}} = \hat{\mathbf{f}} - \mathbf{f}_{\hat{\mathbf{x}}}\hat{\mathbf{v}}$;

Step 3.

Calculate the "weight matrix" $\mathbf{w}$ from $\mathbf{w} = (\mathbf{f}_{\hat{\mathbf{x}}}\boldsymbol{\sigma}\mathbf{f}_{\hat{\mathbf{x}}}^{T})^{-1}$;

Step 4.

Solve the corrections to the parameters, $\boldsymbol{\delta}$, from the linear equations $(\mathbf{f}_{\hat{\mathbf{a}}}^{T}\mathbf{w}\mathbf{f}_{\hat{\mathbf{a}}})\boldsymbol{\delta} = -\mathbf{f}_{\hat{\mathbf{a}}}^{T}\mathbf{w}\hat{\boldsymbol{\phi}}$;

Step 5.

Calculate the improved residuals $\hat{\mathbf{v}}_n$ from

$$\hat{\mathbf{v}}_n = -\boldsymbol{\sigma}\mathbf{f}_{\hat{\mathbf{x}}}^{T}\mathbf{w}(\hat{\boldsymbol{\phi}}+\mathbf{f}_{\hat{\mathbf{a}}}\boldsymbol{\delta}) \quad ;$$

Step 6.

Find the new approximate solution from

$$\hat{\mathbf{a}}_n = \hat{\mathbf{a}} + \boldsymbol{\delta} \quad ,$$

$$\hat{\mathbf{x}}_n = \mathbf{x}_0 + \hat{\mathbf{v}}_n \quad ;$$

Step 7.

Test the relative magnitude of each component of $\boldsymbol{\delta}$ and $\hat{\mathbf{v}}$ against $\hat{\mathbf{a}}$ and $\hat{\mathbf{x}}$, and decide if a further iteration is needed, in which case all the steps above must be repeated.

In detail, the steps are as follows.

Step 1.

Calculate the vector of functions in condition equations $\mathbf{f}$, and the vectors of partial derivatives $\mathbf{f}_{\hat{\mathbf{x}}}$ and $\mathbf{f}_{\hat{\mathbf{a}}}$ from the current values of $\hat{\mathbf{x}}$ and $\hat{\mathbf{a}}$, where $\hat{\mathbf{x}}=\mathbf{x}_0+\hat{\mathbf{v}}$.

The dimension of the vectors $\mathbf{x}_0$, $\hat{\mathbf{v}}$, $\hat{\mathbf{x}}$ all are 2m, m being the number of observed positions.

The observation vector is

$$\mathbf{x}_0 = (x_{10},y_{10},\ldots,x_{i0},y_{i0},\ldots,x_{m0},y_{m0})^T \quad . \tag{4-11}$$

The current vector of corrections to observations is

$$\hat{\mathbf{v}} = (v_{x1},v_{y1},\ldots,v_{xi},v_{yi},\ldots,v_{xm},v_{ym})^T \quad . \tag{4-12}$$

From $\mathbf{x}_0$ and $\hat{\mathbf{v}}$, $\hat{\mathbf{x}}$ can be easily found by $\hat{\mathbf{x}}=\mathbf{x}_0+\hat{\mathbf{v}}$,

$$\hat{\mathbf{x}} = (x_{10}+v_{x1},y_{10}+v_{y1},\ldots,x_{i0}+v_{xi},y_{i0}+v_{y},\ldots \\ \ldots,x_{m0}+v_{xm},y_{m0}+v_{ym})^T \quad . \tag{4-13}$$

The current approximation for the seven parameters, $\hat{\mathbf{a}}$, is

$$\hat{\mathbf{a}} = (n,T,a,e,i,\omega,\Omega)^T \quad , \tag{4-14}$$

at current values. Insert the known $(\hat{\mathbf{x}},\hat{\mathbf{a}})$ into the

condition equations to get the function 2m-vector $\hat{\mathbf{f}}$. It has a form

$$\hat{\mathbf{f}} = [f_{11}, f_{21}, \ldots, f_{1i}, f_{2i}, \ldots, f_{1m}, f_{2m}]^T \quad , \qquad (4\text{-}15)$$

where

$$f_{1i} = \frac{X_i}{a} + e - \cos E_i \quad , \qquad (4\text{-}16a)$$

$$f_{2i} = \frac{Y_i}{b} - \sin E_i \quad , \qquad (4\text{-}16b)$$

with

$$b = a\sqrt{1-e^2} \quad , \qquad (4\text{-}16c)$$

and

$$E_i = n(t_i - T) + \frac{eY_i}{b} \quad . \qquad (4\text{-}16d)$$

The coordinates X,Y are calculated from Eqs. (3-9), A,B,F,G,M from Eqs. (3-8).

The first derivatives $\mathbf{f}_{\hat{\mathbf{x}}}$, $\mathbf{f}_{\hat{\mathbf{a}}}$ are calculated at current values of $\hat{\mathbf{x}}$ and $\hat{\mathbf{a}}$.

The partial derivatives of $\hat{\mathbf{f}}$ with respect to the observations $\hat{\mathbf{x}}$, $\mathbf{f}_{\hat{\mathbf{x}}}$, is a block-diagonal 2mx2m square matrix of the form

$$
\mathbf{f}_{\hat{\mathbf{x}}} = \begin{pmatrix}
\frac{\partial f_{11}}{\partial x_1} & \frac{\partial f_{11}}{\partial y_1} & & & & & & 0 & \\
\frac{\partial f_{21}}{\partial x_1} & \frac{\partial f_{21}}{\partial y_1} & & & & & & & \\
 & & \ddots & & & & & & \\
 & & & \frac{\partial f_{1i}}{\partial x_i} & \frac{\partial f_{1i}}{\partial y_i} & & & & \\
 & & & \frac{\partial f_{2i}}{\partial x_i} & \frac{\partial f_{2i}}{\partial y_i} & & & & \\
 & & & & & \ddots & & & \\
 & & 0 & & & & \frac{\partial f_{1m}}{\partial x_m} & \frac{\partial f_{1m}}{\partial y_m} & \\
 & & & & & & \frac{\partial f_{2m}}{\partial x_m} & \frac{\partial f_{2m}}{\partial y_m} &
\end{pmatrix}, \tag{4-17}
$$

i.e.,

$$
\mathbf{f}_{\hat{\mathbf{x}}} = \mathrm{diag}(g_1, g_2, \ldots, g_i, \ldots, g_m) \tag{4-18}
$$

with

$$
g_i = \begin{pmatrix}
\frac{\partial f_{1i}}{\partial x_i} & \frac{\partial f_{1i}}{\partial y_i} \\
\frac{\partial f_{2i}}{\partial x_i} & \frac{\partial f_{2i}}{\partial y_i}
\end{pmatrix} \quad i=1,m \quad . \tag{4-19}
$$

From Eqs. (4-16), (3-8) and (3-9), we obtain

$$g_i = \begin{pmatrix} \frac{1}{a} & \frac{e}{b}\sin E_i \\ 0 & \frac{1}{b}(1-e\cos E_i) \end{pmatrix} \begin{pmatrix} \frac{\partial X_i}{\partial x_i} & \frac{\partial X_i}{\partial y_i} \\ \frac{\partial Y_i}{\partial x_i} & \frac{\partial Y_i}{\partial y_i} \end{pmatrix} . \qquad (4\text{-}20)$$

In particular, we have

$$\frac{\partial X_i}{\partial x_i} = \frac{\partial X}{\partial x} = \frac{G}{M} , \qquad \frac{\partial X_i}{\partial y_i} = \frac{\partial X}{\partial y} = -\frac{F}{M} ,$$

$$\frac{\partial Y_i}{\partial x_i} = \frac{\partial Y}{\partial x} = -\frac{B}{M} , \qquad \frac{\partial Y_i}{\partial y_i} = \frac{\partial Y}{\partial y} = \frac{A}{M} , \qquad (4\text{-}21)$$

and therefore

$$g_i = \begin{bmatrix} \frac{1}{a} & \frac{e}{b}\sin E_i \\ 0 & \frac{1}{b}(1-e\cos E_i) \end{bmatrix} \begin{bmatrix} \frac{G}{M} & -\frac{F}{M} \\ -\frac{B}{M} & \frac{A}{M} \end{bmatrix} . \qquad (4\text{-}22)$$

The expressions for all the partial derivatives in $\mathbf{f}_{\hat{\mathbf{x}}}$ are listed in Table 4-1.

The dimension of $\mathbf{f}_{\hat{\mathbf{a}}}$, the vector of the partial derivatives of $\mathbf{f}$ with respect to the seven parameters, is 2mx7. It has the form

$$\mathbf{f}_{\hat{\mathbf{a}}} = (G_1, G_2, \ldots, G_i, \ldots, G_m)^T , \qquad (4\text{-}23)$$

Table 4-1.

Expressions for All the Partial Derivatives in $\mathbf{f}_{\hat{\mathbf{x}}}$

	f_{1i}	f_{2i}
$\frac{\partial}{\partial x}$	$\frac{1}{M}\left(\frac{G}{a} - \frac{B}{b} e\sin E_i\right)$	$-\frac{1}{M}\frac{B}{b}(1-e\cos E_i)$
$\frac{\partial}{\partial y}$	$-\frac{1}{M}\left(\frac{F}{a} - \frac{A}{b} e\sin E_i\right)$	$\frac{1}{M}\frac{A}{b}(1-e\cos E_i)$

where

$$G_i = \begin{pmatrix} \frac{\partial f_{1i}}{\partial n} & \frac{\partial f_{1i}}{\partial T} & \frac{\partial f_{1i}}{\partial a} & \frac{\partial f_{1i}}{\partial e} & \frac{\partial f_{1i}}{\partial i} & \frac{\partial f_{1i}}{\partial \omega} & \frac{\partial f_{1i}}{\partial \Omega} \\ \frac{\partial f_{2i}}{\partial n} & \frac{\partial f_{2i}}{\partial T} & \frac{\partial f_{2i}}{\partial a} & \frac{\partial f_{2i}}{\partial e} & \frac{\partial f_{2i}}{\partial i} & \frac{\partial f_{2i}}{\partial \omega} & \frac{\partial f_{2i}}{\partial \Omega} \end{pmatrix} .$$

The expressions for all the elements in G_i are listed below.

$$\begin{pmatrix} \frac{\partial f_{1i}}{\partial n} & \frac{\partial f_{1i}}{\partial T} \\ \frac{\partial f_{2i}}{\partial n} & \frac{\partial f_{2i}}{\partial T} \end{pmatrix} = \begin{pmatrix} \sin E_i \\ -\cos E_i \end{pmatrix} \begin{pmatrix} t_i - T & -n \end{pmatrix} ; \tag{4-24a}$$

$$\frac{\partial f_{1i}}{\partial a} = -\frac{X_i}{a^2} - \frac{eY_i}{ab}\sin E_i ; \tag{4-24b}$$

$$\frac{\partial f_{2i}}{\partial a} = -\frac{Y_i}{ab}(1 - e\cos E_i) ; \tag{4-24c}$$

$$\frac{\partial f_{1i}}{\partial e} = \frac{Y_i \sin E_i}{b(1-e^2)} + 1 ; \tag{4-24d}$$

$$\frac{\partial f_{2i}}{\partial e} = \frac{e - \cos E_i}{b(1-e^2)} Y_i ; \tag{4-24e}$$

$$\begin{pmatrix} \frac{\partial f_{1i}}{\partial i} & \frac{\partial f_{1i}}{\partial \omega} & \frac{\partial f_{1i}}{\partial \Omega} \\ \frac{\partial f_{2i}}{\partial i} & \frac{\partial f_{2i}}{\partial \omega} & \frac{\partial f_{2i}}{\partial \Omega} \end{pmatrix} = \begin{pmatrix} \frac{1}{a} & \frac{e}{b}\sin E_i \\ 0 & \frac{1}{b}(1-e\cos E_i) \end{pmatrix} \begin{pmatrix} \frac{\partial X_i}{\partial i} & \frac{\partial X_i}{\partial \omega} & \frac{\partial X_i}{\partial \Omega} \\ \frac{\partial Y_i}{\partial i} & \frac{\partial Y_i}{\partial \omega} & \frac{\partial Y_i}{\partial \Omega} \end{pmatrix} . \tag{4-24f}$$

From Eqs. (3-8) and (3-9) we can find the expressions for the following six partial derivatives.

$$M\frac{\partial X_i}{\partial i} = z_i \sin\omega \quad , \qquad M\frac{\partial Y_i}{\partial i} = z_i \cos\omega \quad ;$$

$$M\frac{\partial X_i}{\partial \omega} = Ay_i - Bx_i \quad , \quad M\frac{\partial Y_i}{\partial \omega} = Fy_i - Gx_i \quad ;$$

$$M\frac{\partial X_i}{\partial \Omega} = Gy_i + Fx_i \quad , \quad M\frac{\partial Y_i}{\partial \Omega} = -By_i - Ax_i \quad . \tag{4-24g}$$

In terms of these derivatives, we have

$$\begin{pmatrix} \frac{\partial f_{1i}}{\partial i} & \frac{\partial f_{1i}}{\partial \omega} & \frac{\partial f_{1i}}{\partial \Omega} \\ \frac{\partial f_{2i}}{\partial i} & \frac{\partial f_{2i}}{\partial \omega} & \frac{\partial f_{2i}}{\partial \Omega} \end{pmatrix} = \begin{pmatrix} \frac{1}{Ma} & \frac{e}{Mb}\sin E_i \\ 0 & \frac{1}{Mb} - \frac{e}{Mb}\cos E_i \end{pmatrix} \begin{pmatrix} z_i \sin\omega & Ay_i - Bx_i & Gy_i + Fx_i \\ z_i \cos\omega & Fy_i - Gx_i & -By_i - Ax_i \end{pmatrix} . \tag{4-24h}$$

All expressions in $f_{\hat{a}}$ are listed in Table 4-2.

Table 4-2.

Expressions for All the Partial Derivatives in $f_{\hat{a}}$

	f_{1i}	f_{2i}
$\frac{\partial}{\partial n}$	$(t_i-T)\sin E_i$	$-(t_i-T)\cos E_i$
$\frac{\partial}{\partial T}$	$-n\sin E_i$	$n\cos E_i$
$\frac{\partial}{\partial a}$	$-\frac{X_i}{a^2}-\frac{Y_i}{ab}e\sin E_i$	$-\frac{Y_i}{ab}(1-e\cos E_i)$
$\frac{\partial}{\partial e}$	$1+\frac{Y_i}{b(1-e^2)}\sin E_i$	$\frac{Y_i}{b(1-e^2)}(e-\cos E_i)$
$\frac{\partial}{\partial i}$	$\frac{z_i}{M}\left[\frac{\sin\omega}{a}+\frac{e}{b}\sin E_i\cos\omega\right]$	$\frac{1}{Mb}(1-e\cos E_i)z_i\cos\omega$
$\frac{\partial}{\partial \omega}$	$\frac{1}{M}\left[\frac{Ay_i-Bx_i}{a}+\frac{e}{b}\sin E_i(Fy_i-Gx_i)\right]$	$\frac{1}{Mb}(1-e\cos E_i)(Fy_i-Gx_i)$
$\frac{\partial}{\partial \Omega}$	$\frac{1}{M}\left[\frac{Gy_i+Fx_i}{a}-\frac{e}{b}\sin E_i(By_i+Ax_i)\right]$	$\frac{1}{Mb}(e\cos E_i-1)(By_i+Ax_i)$

Step 2.

Calculate $\hat{\phi}$ from $\mathbf{f}$, $\hat{\mathbf{v}}$ and $\mathbf{f}_{\hat{\mathbf{x}}}$ by $\hat{\phi}=\mathbf{f}-\mathbf{f}_{\hat{\mathbf{x}}}\hat{\mathbf{v}}$. The dimension of the vector $\hat{\phi}$ is also 2m. It has the form

$$\hat{\phi} = (\phi_1,\phi_2,\ldots,\phi_i,\ldots,\phi_m)^T \quad , \tag{4-25}$$

where

$$\phi_i = \begin{pmatrix} f_{1i} - \dfrac{\partial f_{1i}}{\partial x_i} v_{xi} - \dfrac{\partial f_{1i}}{\partial y_i} v_{yi} \\ \\ f_{2i} - \dfrac{\partial f_{2i}}{\partial x_i} v_{xi} - \dfrac{\partial f_{2i}}{\partial y_i} v_{yi} \end{pmatrix} = \begin{pmatrix} \phi_{i1} \\ \\ \phi_{i2} \end{pmatrix} \tag{4-25'}$$

Step 3.

Calculate the weight matrix $\mathbf{w}$ from $\mathbf{f}_{\hat{\mathbf{x}}}$ and σ.

The matrix $\mathbf{f}_{\hat{\mathbf{x}}}$ has been calculated in step 1. The covariance matrix σ is assumed to be known. The dimension of σ is 2mx2m.

An example for computing σ is shown below.

The relationship between the covariance matrix of rectangular coordinates (x, y) and that of polar coordinates (ρ, θ) is

$$\sigma_{xy} = \left[\frac{\partial(x,y)}{\partial(\rho,\theta)} \right] \sigma_{\rho\theta} \left[\frac{\partial(x,y)}{\partial(\rho,\theta)} \right]^T , \tag{4-26}$$

where $x=\rho\cos\theta$, $y=\rho\sin\theta$ and

$$\frac{\partial(x,y)}{\partial(\rho,\theta)} = \begin{pmatrix} \cos\theta & -\rho\sin\theta \\ \sin\theta & \rho\cos\theta \end{pmatrix} . \qquad (4\text{-}26')$$

Thus, we have

$$\sigma_{xy} = \begin{pmatrix} \cos\theta & -\rho\sin\theta \\ \sin\theta & \rho\cos\theta \end{pmatrix} \begin{pmatrix} \sigma_\rho & 0 \\ 0 & \sigma_\theta \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\rho\sin\theta & \rho\cos\theta \end{pmatrix}$$

and therefore

$$\sigma_{xy} = \begin{pmatrix} \sigma_\rho\cos^2\theta + \sigma_\theta\rho^2\sin^2\theta & (\sigma_\rho - \sigma_\theta\rho^2)\cos\theta\sin\theta \\ (\sigma_\rho - \sigma_\theta\rho^2)\cos\theta\sin\theta & \sigma_\rho\sin^2\theta + \sigma_\theta\rho^2\cos^2\theta \end{pmatrix} . \qquad (4\text{-}26'')$$

In these expressions, $\sigma_\rho=\Delta^2\rho$, $\sigma_\theta=\Delta^2\theta$ and $\Delta\rho$, $\Delta\theta$ are the observational errors in ρ and θ, respectively. For the observations, the random errors are of similar order of magnitude as the systematic ones, larger in separation than in position angle. The average errors $\rho\Delta\theta$ and $\Delta\rho$ vary somewhat with the separation ρ and can be assumed, for many series of observations, to follow the form $C\rho^{1/3}$, where C varies with different observers. For a single good observation C will not exceed 0".03 in position angle ($\rho\Delta\theta$) and 0".08 in separation ($\Delta\rho$) (Heintz, 1971). Errors will be somewhat larger and difficult to measure if one or both components are faint. If the errors are expressed in the

dimensionless (relative) forms $\Delta\theta$ and $\Delta\rho/\rho$, it is seen that they increase as pairs become closer.

Based on the considerations above, we can, for example, put $\Delta\rho=0''.08\rho^{1/3}$ and $\rho\Delta\theta=0''.03\rho^{1/3}$, i.e. $\Delta\theta=0''.03\rho^{-2/3}$. If we allow each pair of numbers (x_i,y_i) in an observation to be correlated, but no correlations between different observations, σ, would be block-diagonal. In this case, we have

$$\sigma = \mathrm{diag}(\sigma_1,\sigma_2,\ldots,\sigma_i,\ldots,\sigma_m) \quad ,$$

where

$$\sigma_i = \begin{pmatrix} \sigma_{xixi} & \sigma_{xiyi} \\ \sigma_{yixi} & \sigma_{yiyi} \end{pmatrix} = \begin{pmatrix} \sigma_{i1} & \sigma_{i3} \\ \sigma_{i4} & \sigma_{i2} \end{pmatrix} \quad ,$$

with $\sigma_{xiyi}=\sigma_{yixi}$, i.e., $\sigma_{i3}=\sigma_{i4}$.

The form of the weight matrix is simpler in this case; it is also block-diagonal.

According to Eq. (4-5), $\mathbf{w}=(\mathbf{f}_{\hat{\mathbf{x}}}\sigma\mathbf{f}_{\hat{\mathbf{x}}}^{\mathrm{T}})^{-1}$. The matrices $\mathbf{f}_{\hat{\mathbf{x}}}$, σ, $\mathbf{f}_{\hat{\mathbf{x}}}^{\mathbf{T}}$ now are all block-diagonal. Therefore

$$\mathbf{w} = \mathrm{diag}(w_1,w_2,\ldots,w_i,\ldots,w_m) \quad , \qquad (4\text{-}29)$$

with

$$w_i = \begin{pmatrix} w_{i1} & w_{i3} \\ w_{i4} & w_{i2} \end{pmatrix} \quad . \tag{4-30}$$

The computation of $\mathbf{w}$ is straightforward. We first find $\mathbf{w}^{-1}$. If we denote $\mathbf{u}=\mathbf{w}^{-1}=\mathbf{f}_{\hat{\mathbf{x}}}\boldsymbol{\sigma}\mathbf{f}_{\hat{\mathbf{x}}}^{\mathbf{T}}$, it is obvious that $\mathbf{u}$ has the same form

$$\mathbf{u} = \mathrm{diag}(u_1, u_2, \ldots, u_i, \ldots, u_m) \quad , \tag{4-31}$$

where

$$u_i = \begin{pmatrix} u_{i1} & u_{i3} \\ u_{i4} & u_{i2} \end{pmatrix} =$$

$$= \begin{pmatrix} \dfrac{\partial f_{1i}}{\partial x_i} & \dfrac{\partial f_{1i}}{\partial y_i} \\ \dfrac{\partial f_{2i}}{\partial x_i} & \dfrac{\partial f_{2i}}{\partial y_i} \end{pmatrix} \begin{pmatrix} \sigma_{i1} & \sigma_{i3} \\ \sigma_{i4} & \sigma_{i2} \end{pmatrix} \begin{pmatrix} \dfrac{\partial f_{1i}}{\partial x_i} & \dfrac{\partial f_{2i}}{\partial x_i} \\ \dfrac{\partial f_{1i}}{\partial y_i} & \dfrac{\partial f_{2i}}{\partial y_i} \end{pmatrix} , \tag{4-32}$$

i.e.,

$$u_{i1} = \left(\frac{\partial f_{1i}}{\partial x_i}\right)^2 \sigma_{i1} + 2 \frac{\partial f_{1i}}{\partial x_i} \frac{\partial f_{1i}}{\partial y_i} \sigma_{i3} + \left(\frac{\partial f_{1i}}{\partial y_i}\right)^2 \sigma_{i2} \quad ; \tag{4-33a}$$

$$u_{i2} = \left(\frac{\partial f_{2i}}{\partial x_i}\right)^2 \sigma_{i1} + 2 \frac{\partial f_{2i}}{\partial x_i} \frac{\partial f_{2i}}{\partial y_i} \sigma_{i3} + \left(\frac{\partial f_{2i}}{\partial y_i}\right)^2 \sigma_{i2} \quad ; \qquad (4\text{-}33b)$$

$$u_{i3} = \frac{\partial f_{1i}}{\partial x_i} \frac{\partial f_{2i}}{\partial x_i} \sigma_{i1} + \left(\frac{\partial f_{1i}}{\partial x_i} \frac{\partial f_{2i}}{\partial y_i} + \frac{\partial f_{1i}}{\partial y_i} \frac{\partial f_{2i}}{\partial x_i}\right) \sigma_{i3} +$$

$$\frac{\partial f_{1i}}{\partial y_i} \frac{\partial f_{2i}}{\partial y_i} \sigma_{i2} \quad ; \qquad (4\text{-}33c)$$

$$u_{i4} = u_{i3} \quad . \qquad (4\text{-}33d)$$

After computing **u**, we can find its inverse **w** very easily. If we denote $u=u_{i1}u_{i2}-u_{i3}u_{i4}=u_{i1}u_{i2}-u_{i3}^2$, we have

$$w_i = \begin{pmatrix} w_{i1} & w_{i3} \\ w_{i4} & w_{i2} \end{pmatrix} = \begin{pmatrix} \frac{u_{i2}}{u} & -\frac{u_{i4}}{u} \\ -\frac{u_{i3}}{u} & \frac{u_{i1}}{u} \end{pmatrix} \quad . \qquad (4\text{-}34)$$

As seen above, we can calculate the components of **w** directly from the components of **u**.

If the quantities which constitute different observations were also correlated, we would have to find the product of the matrices $\mathbf{f}_{\tilde{\mathbf{X}}}$, σ, $\mathbf{f}_{\tilde{\mathbf{X}}}^{\mathbf{T}}$ and calculate its inverse to get **w**. This would be more complicated by far.

Step 4.

Find $\mathbf{f}_{\hat{a}}^T \mathbf{w} \mathbf{f}_{\hat{a}}$ and $-\mathbf{f}_{\hat{a}}^T \hat{\boldsymbol{\phi}}$ from $\mathbf{f}_{\hat{a}}$, $\hat{\boldsymbol{\phi}}$ and $\mathbf{w}$, then solve the linear equations

$$(\mathbf{f}_{\hat{a}}^T \mathbf{w} \mathbf{f}_{\hat{a}})\boldsymbol{\delta} = \mathbf{f}_{\hat{a}}^T \mathbf{w} \hat{\boldsymbol{\phi}}$$

to get the correction vector $\boldsymbol{\delta}$.

The dimension of the matrix $\mathbf{f}_{\hat{a}}^T \mathbf{w} \mathbf{f}_{\hat{a}}$ is 7x7 and $-\mathbf{f}_{\hat{a}}^T \mathbf{w} \hat{\boldsymbol{\phi}}$, $\boldsymbol{\delta}$ are 7-vectors. Denoting $\mathbf{A}=\mathbf{f}_{\hat{a}}^T \mathbf{w} \mathbf{f}_{\hat{a}}$ and $\mathbf{B}=\mathbf{f}_{\hat{a}}^T \mathbf{w} \hat{\boldsymbol{\phi}}$, we have

$$A(j,k) = \sum_{i=1}^{m} \left(\frac{\partial f_{1i}}{\partial a_j} \frac{\partial f_{1i}}{\partial a_k} w_{i1} + \frac{\partial f_{2i}}{\partial a_j} \frac{\partial f_{1i}}{\partial a_k} w_{i3} + \right.$$

$$\left. + \frac{\partial f_{1i}}{\partial a_j} \frac{\partial f_{2i}}{\partial a_k} w_{i3} + \frac{\partial f_{2i}}{\partial a_j} \frac{\partial f_{2i}}{\partial a_k} w_{i2} \right)$$

$$j=1,7;$$

$$k=1,7; \quad (4\text{-}35a)$$

and

$$B(j) = \sum_{i=1}^{m} \left(\frac{\partial f_{1i}}{\partial a_j} w_{i1} \phi_{i1} + \frac{\partial f_{2i}}{\partial a_j} w_{i3} \phi_{i1} + \right.$$

$$\left. + \frac{\partial f_{1i}}{\partial a_j} w_{i3} \phi_{i2} + \frac{\partial f_{2i}}{\partial a_j} w_{i2} \phi_{i2} \right)$$

$$j=1,7; \quad (4\text{-}35b)$$

where

$$\frac{\partial f_{1i}}{\partial a_1} = \frac{\partial f_{1i}}{\partial n}, \quad \frac{\partial f_{1i}}{\partial a_2} = \frac{\partial f_{1i}}{\partial T}, \quad \frac{\partial f_{1i}}{\partial a_3} = \frac{\partial f_{1i}}{\partial a}, \quad \frac{\partial f_{1i}}{\partial a_4} = \frac{\partial f_{1i}}{\partial e},$$

..., and so forth.

To solve the linear equations $\mathbf{A}\delta=-\mathbf{B}$ for δ, we can apply any linear equation solution method, e.g. Gaussian Elimination, Gauss-Jordan Elimination, the Jacobi Iterative method, etc. In a private communication, Eichhorn (1985) proposed a special solution method for this particular problem, because the covariance matrix, i.e., the inverse of the coefficient matrix of δ, is fortunately a positive definite matrix. The method will be described in another section of this chapter. In our work, we use this special method for solving the linear equations above for δ.

Step 5.

Calculate the new vector $\hat{\mathbf{v}}_n$, which now is the improved correction vector to the observations, from

$$\hat{\mathbf{v}}_n = - \sigma \mathbf{f}_{\hat{\mathbf{x}}}^T \mathbf{w} (\hat{\phi} + \mathbf{f}_{\hat{\mathbf{a}}} \delta) \quad .$$

The dimension of the vector $\hat{\mathbf{v}}$ is 2m. Denote $\mathbf{Q}=\hat{\phi}\mathbf{f}_{\hat{\mathbf{a}}}\delta$ and $\mathbf{R}=\mathbf{f}_{\hat{\mathbf{x}}}^T\mathbf{w}\mathbf{Q}$, $\mathbf{Q}$ and $\mathbf{R}$ both are also 2m-vectors. We can see that

$$\mathbf{Q} = (Q_1, Q_2, \ldots, Q_i, \ldots, Q_m)^T \quad , \qquad (4\text{-}36)$$

where

$$Q_i = \begin{pmatrix} f_{1i} - \dfrac{\partial f_{1i}}{\partial x_i} v_{xi} - \dfrac{\partial f_{1i}}{\partial y_i} v_{yi} + \sum_{i=1}^{7} \dfrac{\partial f_{1i}}{\partial a_j} \delta_j \\ \\ f_{2i} - \dfrac{\partial f_{2i}}{\partial x_i} v_{xi} - \dfrac{\partial f_{2i}}{\partial y_i} v_{yi} + \sum_{i=1}^{7} \dfrac{\partial f_{2i}}{\partial a_j} \delta_j \end{pmatrix} ,$$

i.e.,

$$Q_i = \begin{pmatrix} f_{1i} + \sum_{i=1}^{7} \dfrac{\partial f_{1i}}{\partial a_j} \delta_j \\ \\ f_{2i} + \sum_{i=1}^{7} \dfrac{\partial f_{2i}}{\partial a_j} \delta_j \end{pmatrix} = \begin{pmatrix} Q_{i1} \\ \\ Q_{i2} \end{pmatrix} . \qquad (4\text{-}36')$$

In the case of **w** being block-diagonal,

$$\mathbf{R} = (R_1, R_2, \ldots, R_i, \ldots, R_m)^T , \qquad (4\text{-}37)$$

where

$$R_i = \begin{pmatrix} R_{i1} \\ \\ R_{i2} \end{pmatrix} =$$

$$= \begin{pmatrix} \dfrac{\partial f_{1i}}{\partial x_i}\,(Q_{i1}W_{i1}+Q_{i2}W_{i3}) + \dfrac{\partial f_{2i}}{\partial x_i}\,(Q_{i1}W_{i4}+Q_{i2}W_{i2}) \\ \\ \dfrac{\partial f_{1i}}{\partial y_i}\,(Q_{i1}W_{i1}+Q_{i2}W_{i3}) + \dfrac{\partial f_{2i}}{\partial y_i}\,(Q_{i1}W_{i4}+Q_{i2}W_{i2}) \end{pmatrix} . \qquad (4\text{-}37')$$

Therefore, in this case,

$$\hat{\mathbf{v}}_n = (V_1, V_2, \ldots, V_i, \ldots, V_m)^T \quad , \qquad (4\text{-}38)$$

where

$$V_i = \begin{pmatrix} V_{xi} \\ \\ V_{yi} \end{pmatrix} = \begin{pmatrix} \sigma_{i1}R_{i1} + \sigma_{i3}R_{i2} \\ \\ \sigma_{i4}R_{i1} + \sigma_{i2}R_{i2} \end{pmatrix} . \qquad (4\text{-}38')$$

Step 6.

Find the new approximate solutions $\hat{\mathbf{a}}_n$ and $\hat{\mathbf{x}}_n$ from

$$\hat{\mathbf{a}}_n = \hat{\mathbf{a}} + \boldsymbol{\delta} \quad ,$$
$$\hat{\mathbf{x}}_n = \hat{\mathbf{x}}_0 + \hat{\mathbf{v}}_n \quad .$$

Step 7.

Determine if another iteration is needed.

Calculate all quantities of

$$\left| \frac{x_{i(new)} - x_{i(old)}}{x_{i(old)}} \right| , \left| \frac{y_{i(new)} - y_{i(old)}}{y_{i(old)}} \right| \text{ and } \left| \frac{\delta_j}{a_j} \right|$$

(altogether 2m+7 quantities) and pick up the maximum value among them. If this maximum exceeds a pre-set specified value, say 10^{-8}, return to the very beginning and take all of the steps once again. Only when the maximum is less than the specified value, we assume that good convergence has been achieved.

Above is the scheme of Newton's method in our orbit problem. In real programming, some additional considerations have been taken into account. For example, in the actual program, all variables are dimensionless. For rectangular coordinate pairs (x, y), two new quantities are defined,

$$x' = \frac{x}{x_0} , \qquad y' = \frac{y}{y_0} , \tag{4-39}$$

where (x_0, y_0) are the "observations" in the initial data, so that $x = x_0 x'$, $y = y_0 y'$. For the seven parameters, seven new variables are defined as well. They are

$$a_1 = \frac{n}{n_0} , \quad a_2 = \frac{T}{T_0} , \quad a_3 = \frac{a}{a_0} , \quad a_4 = \frac{e}{e_0} , \quad a_5 = \frac{i}{i_0} ,$$

$$a_6 = \frac{\omega}{\omega_0} , \quad a_7 = \frac{\Omega}{\Omega_0} , \tag{4-39'}$$

where $(n_0, T_0, a_0, e_0, i_0, \omega_0, \Omega_0)$ in $\hat{a}_0$ is the initial approximate solution and e is defined as $e=\sin\phi$ and $e_0=\sin\phi_0$.

In terms of these new variables, all formulae for computing the partial derivatives need only slight modifications, and the whole process remains in principle unchanged.

In addition, we have seen that $\mathbf{f}_{\hat{\mathbf{x}}}$, σ, $\mathbf{f}_{\hat{\mathbf{x}}}^{\mathbf{T}}$ are square 2mx2m matrices. If the number of observations is, e.g. m=100, the covariance matrix σ has 40,000 components, and these three matrices alone occupy a huge storage, 120,000, in the computer. In practical programming, we should keep the required computer storage to a possible minimum. A block-diagonal covariance matrix σ can be stored in a 2mx2 matrix, and the same applies to all other related matrices. This greatly reduces the need for storage.

Weighting of Observations

As we know, the measures of any binary star will be affected by the accidental and systematic observational errors and, occasionally, from blunders, i.e., actual mistakes. The points, when plotted, will not lie exactly on an ellipse but will occupy a region which is clustered around an ellipse only in a general way.

Observations are frequently combined into normal places. Among the observed quantities, the time is the one observed most precisely. To investigate the measurements for discordance before using them to calculate an orbit, the simplest method is to plot the distance ρ in seconds of arc and the position angles θ, separately, as ordinates, against the times of observation as abscissae. Smooth curves are

drawn to represent the general run of the measures and in drawing these curves, more consideration will naturally be given to those observation points which are relatively more precise (for example, a point based upon several well agreeing measures by a skilled observer and supported by the preceding and following observations) than to the others. The deviation of the observation points is in the ordinate direction only and gives a good idea of the accuracy of both observed quantities. The curves will show whether or not the measures as a whole are sufficiently good to warrant the determination of a reasonably reliable orbit. Observations which are seriously in error will be clearly revealed and should be rejected or given very low weights. The curves will also give a general idea of how the observations are to be weighted. The points which show larger deviations should be given lower weights and the well-determined points should be given higher weights.

It is hard to recommend a general rule for the weighting of measurements and normal positions. Some precept could be considered (W. D. Heintz, 1971): compute a weight p_1 according to the number of observations, and p_2 according to the "weight" assigned to the observers, then the normal place receives the weight $\rho\sqrt{p_1 p_2}$ (if computations are made in the quantities $d\theta$ and $d\rho/\rho$). This precept could avoid unduly high weights for single measurements as well as for very many observations by few observers, and would

reduce the influence of residual systematic errors. Its implicit assumption is a proportionality of the errors $\rho d\theta$ and $d\rho$ with $\sqrt{\rho}$. This holds better for the multi-observer average, as the share by less accurate data usually increases at larger separations.

In our work, we given the points on or nearly on the smooth curves equal unit weights, and assign lower weights to those farther from the lines.

When we take into account the weights for the observations, some slight and very simple modification is needed in the solution process described above.

Let **G** be the matrix of weights for the observational errors $\hat{\mathbf{v}}$. The dimension of **G** is 2mx2m, and **G** is diagonal.

Taking into account **G**, the residual function S_0 will be modified as

$$S_0 = \frac{1}{2}\hat{\mathbf{v}}^T\mathbf{G}^T\sigma^{-1}\mathbf{G}\hat{\mathbf{v}} = \frac{1}{2}\hat{\mathbf{v}}^T\mathbf{G}\sigma^{-1}\mathbf{G}\hat{\mathbf{v}} \quad , \tag{4-40}$$

because $\mathbf{G}^T = \mathbf{G}$. If we define $\Sigma^{-1} = \mathbf{G}\sigma^{-1}\mathbf{G}$, then

$$S_0 = \frac{1}{2}\hat{\mathbf{v}}^T\Sigma^{-1}\hat{\mathbf{v}} \quad , \tag{4-40'}$$

which has the exact form as before except σ is replaced by Σ. Therefore we need only to replace σ by Σ, σ^{-1} by Σ^{-1} wherever σ of σ^{-1} appears.

Because **G** is diagonal, the calculations of Σ and Σ^{-1} from **G** and σ are also very simple.

The Orthogonal Set of Adjustment Parameters and the Efficiency of a Set of Orbital Parameters

As mentioned in the first section of this chapter, Eichhorn (1985) has proposed a special method in a private communication for solving the linear normal equations $\mathbf{A}\delta=-\mathbf{B}$ of our problem. This method is as follows.

According to its definition, $\mathbf{A}=\mathbf{f}_{\hat{a}}^{T}\mathbf{w}\mathbf{f}_{\hat{a}}$ is obviously a symmetrical positive definite matrix.

Given a symmetrical positive definite matrix **Q**, we look for its eigenvectors **X** which have the property that

$$\mathbf{QX} = \lambda\mathbf{X} \quad . \tag{4-41}$$

Any value λ which satisfies Eq. (4-41) is an eigenvalue of **Q**, and since **Q** is positive definite we know that all $\lambda>0$. Obviously λ must satisfy the equation $|\mathbf{Q}-\mathbf{I}\lambda| = 0$, the "characteristic equation" of **Q**. This is a polynomial in λ of the same order n as that of **Q**. The solutions are $\lambda_1, \lambda_2,\ldots,\lambda_n$. The solution $\mathbf{X}_k$ of the homogeneous system (4-41) with $\lambda=\lambda_K$ is the k-th eigenvector. Since this is determined only with the uncertainty of an arbitrary scale factor, we may always achieve $|\mathbf{X}_K| = 1$ for all k. Let the matrix $\mathbf{P}=(\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n)$ be the matrix of the normalized eigenvectors of **Q**. It can be shown that **P** is therefore

orthogonal, so that $\mathbf{P}^T \equiv \mathbf{P}^{-1}$. Equation (4-41) can be written as

$$\mathbf{QP} = \mathbf{P}\,\mathrm{diag}(\lambda_1, \ldots, \lambda_n) \quad . \tag{4-42}$$

Writing the diagonal matrix $\mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) = \mathbf{D}$, we rewrite Eq. (4-42) as

$$\mathbf{QP} = \mathbf{PD} \quad ,$$

whence

$$\mathbf{P}^T\mathbf{QP} = \mathbf{D} \quad , \qquad \mathbf{Q} = \mathbf{PDP}^T \quad , \tag{4-43}$$

and

$$\mathbf{P}^T\mathbf{Q}^{-1}\mathbf{P} = \mathbf{D}^{-1} \quad , \quad \mathbf{Q}^{-1} = \mathbf{PD}^{-1}\mathbf{P}^T \quad . \tag{4-44}$$

The equation (4-44) shows incidentally that the eigenvectors of a matrix are identical to those of its inverse, but that the eigenvalues of a matrix are the inverses of the eigenvalues of the inverse.

Let $\mathbf{Q}$ be the covariance matrix of a set of statistical variates $\mathbf{X}$. We look for another set $\mathbf{Y}$ as function of $\mathbf{X}$, such that the covariance matrix of $\mathbf{Y}$ is diagonal.

Putting $d\mathbf{X}=\boldsymbol{\alpha}$, the quadratic form which is minimized is $\boldsymbol{\alpha}^T\mathbf{Q}^{-1}\boldsymbol{\alpha}$, where $\mathbf{Q}$ is the covariance matrix of $\mathbf{X}$. If we define

$$\mathbf{Y} = \mathbf{P}^T\mathbf{X} \quad , \tag{4-45}$$

then we have

$$\mathbf{X} = \mathbf{PY}$$

and

$$d\mathbf{X} = \mathbf{P}d\mathbf{Y} \quad \text{or} \quad \boldsymbol{\alpha} = \mathbf{P}\boldsymbol{\beta} \ , \quad \boldsymbol{\alpha}^T = \boldsymbol{\beta}^T\mathbf{P}^T \ .$$

and in terms of $\boldsymbol{\beta} = d\mathbf{Y}$, the quadratic form minimized which led to the values $\mathbf{X}$ becomes $\boldsymbol{\beta}^T\mathbf{P}^T\mathbf{Q}^{-1}\mathbf{P}\boldsymbol{\beta}$, whence the covariance matrix of $\mathbf{Y}$ is seen to be $(\mathbf{P}^T\mathbf{Q}^{-1}\mathbf{P})^{-1}\equiv\mathbf{P}^T\mathbf{Q}\mathbf{P}\equiv\mathbf{D}$ by Eq. (4-44).

It is then shown that $\mathbf{Y}=\mathbf{P}^T\mathbf{X}$ is the vector of linear transforms of $\mathbf{X}$ whose components are uncorrelated. A vector of such components is called a vector of statistical variates with efficiency 1 (Eichhorn and Cole, 1985). If any of its components are changed (e.g. in the process of finding their values in a course of iteration), none of the other components of $\mathbf{Y}$ would thereby be changed in a tradeoff between changes of correlated unknowns.

Now, back to the normal equations $\mathbf{A}\delta=-\mathbf{B}$. The problem of solving normal equations above can therefore be attacked as follows.

1) Find the eigenvalues and eigenvectors of $\mathbf{A}$, i.e., $\mathbf{D}^{-1}$ and $\mathbf{P}$. ($\mathbf{A}=\mathbf{Q}^{-1}$, according to the notation above.)

2) The covariance matrix of δ, $\mathbf{Q}$, can then be calculated from Eq. (4-43), $\mathbf{Q}=\mathbf{A}^{-1}=\mathbf{P}\mathbf{D}\mathbf{P}^{T}$.

3) Let $\delta'=\mathbf{P}^{T}\delta$. This gives $\mathbf{A}\mathbf{P}\delta'=-\mathbf{B}$, whence $\delta'=-\mathbf{P}^{T}\mathbf{Q}\mathbf{B}$, or from Eq. (4-43),

$$\delta' = - \mathbf{D}\mathbf{P}^{T}\mathbf{B} \quad . \tag{4-46}$$

Finding $\mathbf{D}^{-1}$ from $\mathbf{D}$ is very easy because $\mathbf{D}$ is diagonal. The vector δ is then easily calculated directly from $\delta=\mathbf{P}\delta'$.

One of the advantages of this method is that we can calculate both δ and δ', the correlated and uncorrelated elements of the corrections to the adjustment parameters, at the same time. The vector δ' is the set of the corrections to the orthogonal adjustment parameters.

Furthermore, using this method, a measure for the "efficiency" of a set of adjustment parameters can be easily calculated, cf. Eichhorn and Cole (1985). They point out that the information carried by the vector δ of the correlated estimates (whose covariance matrix is $\mathbf{Q}$) is partly redundant because of the nonvanishing correlations between the estimates. What is the information content

carried by these estimates? We see that the matrix $\mathbf{P}^T\mathbf{Q}\mathbf{P}$ is diagonal if $\mathbf{P}$ is the (orthogonal) matrix of the normalized eigenvectors of $\mathbf{Q}$ and that it is also the (therefore uncorrelated) linear transforms $\boldsymbol{\delta}'=\mathbf{P}\boldsymbol{\delta}$ of $\boldsymbol{\delta}$. We might regard the number

$$é = \frac{|\mathbf{Q}|}{q_{11}\cdots q_{nn}} , \tag{4-47}$$

that is, the product of the variances of the components of vector $\boldsymbol{\delta}'$ (the product of the variances of the uncorrelated parameters) divided by the product of the variances of the components of vector $\boldsymbol{\delta}$ (the product of the variances of the correlated parameters) as a measure for the "efficiency" of the information carried by the estimates $\boldsymbol{\delta}$. But we note that according to the definition (4-47), the value of é is severely affected by the number of variables. In order to eliminate the effect of n, we redefine é as

$$é = \left(\frac{|\mathbf{Q}|}{q_{11}\cdots q_{nn}} \right)^{\frac{1}{n}} . \tag{4-47'}$$

For any set of uncorrelated estimates we would have e=1, which is evidently the largest value this number may assume.

In our work, for every model calculated, the efficiency of the set of adjustment parameters and their covariance matrix are calculated.

A Practical Example

Table 4-3 lists the observation data for 51 Tau. These data are provided by H. A. Macalister. The author would like to thank Macalister and Heintz for their data which he has used in this dissertation.

Theoretically, the first step in our computation should be the reduction of the measured coordinates to a common epoch by the application to the position angles of corrections for precession and for the proper motion of the system. The distance measures need no corrections. Practically, both corrections are negligibly small unless the star is near the Pole, its proper motion unusually large, and the time covered by the observations long. The precession correction, when required, can be found with sufficient accuracy from the approximate formula

$$\Delta\theta = \theta - \theta_0 = +0^{\circ}00557 \sin\alpha \sec\delta (t-t_0) \quad , \qquad (4\text{-}48)$$

which is derived by differentiating Eq. (3-1) with respect to θ, and introducing the precessional change $\Delta(n\cos\alpha)$ for $d\delta$. The position angles are thus reduced to a common equinox t_0 (for which 2000.0 is currently used), and the resulting node Ω_0 also refers to t_0, because $\Delta\theta=\theta-\theta_0=\Omega-\Omega_0=\Delta\Omega$. Computing ephemerids, the equinox is reduced back from t_0 to t.

The change of position angle by proper motion,

$$\theta - \theta_0 = \mu_\alpha \sin\alpha (t-t_0) \quad , \tag{4-49}$$

where the proper motion component in right ascension μ_α is in degrees, can be neglected in most cases.

The two formulae above can be found either in Heintz' book "Double Stars" or in Aitken's book "The Binary Stars".

In table 4-4, all position angles have been reduced to the common equinox 2000.0 and the converted rectangular coordinates (x_0, y_0) are also listed.

In figure 4-1, all pairs of rectangular coordinates (x_0, y_0) are plotted in the x_0-y_0 plane. We can see at a glance that all observation points are distributed closely to an ellipse. Furthermore, in Figures 2a and 2b, we plot the distance ρ in seconds of arc and the position angles against the observing epoch, respectively. From Figure 2, we see that the observations fall upon nearly sine-like curves. Thus, we get the impression that these data are rather precise and therefore give all observations equal weights.

For these data, an initial approximate solution is easily obtained by Kowalsky's method. The initial approximation $\hat{a}_0$ is listed in Table 4-5.

Starting from this, the final solution for $\hat{a}$ is obtained after only three iterations and shown in Table 4-6. The calculation required only $47^{s}.62$ CPU time on a VAX.

The residuals (observational errors) in (ρ,θ) and (x,y) are shown in Table 4-7. The residuals in (ρ,θ) and (x,y) are plotted against the observing epoch in Figures 4-3 and 4-4, respectively. They show a random distribution as expected. Also, Figure 4-5 shows the comparison of the observation points with the corresponding points after correction in the apparent plane.

The "efficiency," the covariance matrix, the correlation matrix and transformation matrix (which transforms the correlated parameters to the uncorrelated ones) of the adjusted parameters in the final solution are calculated and listed in Table 4-6. In addition, Table 4-6 also lists the standard deviations of a) the original and b) the uncorrelated parameters in the final solution.

Table 4-3.

The Observation Data for 51 Tau.

HR1331	51 Tau	HD 27176	SAO	76541	04185+2135 n
1975.7160	106.0°	2.0	0.080	0.003	A1
1975.9591	91.9	1.7	0.074	0.003	A1
1976.8574	34.9	1.5	0.069	0.005	A2
1976.8602	33.5	1.5	0.073	0.009	A2
1976.9229	22.9	1.0	0.072	0.008	A3
1977.0868	26.7	1.0	0.083	0.008	A3
1977.6398	8.8		0.101		A6
1977.7420	3.1	0.8	0.110	0.008	A5
1978.1490	352.2	0.5	0.113	0.010	A5 n
1978.6183	340.7	0.8	0.108	0.008	A5
1978.8756	333.3	2.0	0.086	0.013	B4
1978.7735	304.3		0.090		A7
1980.1532	285.9		0.075		A8
1980.7182	259.0		0.079		A8
1980.7263	255.8		0.085		A8
1980.7291	259.1		0.087		A8
1982.7550	191.80		0.1343		C2
1982.7579	192.65		0.1362		C2
1982.7605	190.36		0.1315		C2
1982.7633	192.90		0.1381		C2
1982.7661	193.39		0.1308		C2
1983.0472	186.18		0.1333		C2
1983.0637	187.21		0.1499		C2
1983.7108	182.05		0.1456		C2
1983.7135	179.56		0.1480		C2
1983.9337	181.0	1.9	0.149	0.010	FA
1983.9579	176.7		0.157		RB
1984.0522	175.01		0.1446		C2
1984.0576	174.79		0.1445		C2
1984.0603	172.73		0.1355		C2
1984.779	157.3		0.146		RC
1984.9308	164.5	2.6	0.141	0.013	FB
1985.1063	161.0	2.7	0.137	0.013	FB
1985.2048	158.0	3.0	0.125	0.013	FB
1985.8378	145.69		0.1141		##
1985.8406	144.53		0.1202		##
1985.8541	145.71		0.1200		##

Table 4-4.

The Reduced Initial Data for 51 Tau

	t	θ_0	ρ_0	x_0	y_0
1	1975.7160	106.00	0.0800	-0.0222	0.0769
2	1975.9591	91.90	0.0740	-0.0026	0.0740
3	1976.8574	34.90	0.0690	0.0565	0.0396
4	1976.8602	33.50	0.0730	0.0608	0.0404
5	1976.9229	32.90	0.0720	0.0604	0.0392
6	1977.0868	26.70	0.0830	0.0741	0.0374
7	1977.6398	8.80	0.1010	0.0998	0.0156
8	1977.7420	3.10	0.1100	0.1098	0.0061
9	1978.1490	352.20	0.1130	0.1120	-0.0151
10	1978.6183	340.70	0.1080	0.1020	-0.0355
11	1978.8756	333.30	0.0860	0.0769	-0.0385
12	1979.7735	304.30	0.0900	0.0508	-0.0743
13	1980.1532	285.90	0.0750	0.0207	-0.0721
14	1980.7182	259.00	0.0790	-0.0150	-0.0776
15	1980.7263	255.80	0.0850	-0.0207	-0.0824
16	1980.7291	259.10	0.0870	-0.0163	-0.0855
17	1982.7550	191.80	0.1343	-0.1314	-0.0176
18	1982.7579	192.65	0.1362	-0.1329	-0.0300
19	1982.7605	190.36	0.1315	-0.1293	-0.0238
20	1982.7633	192.90	0.1381	-0.1346	-0.0310
21	1982.7661	193.39	0.1308	-0.1272	-0.0305
22	1983.0472	186.18	0.1333	-0.1325	-0.0145
23	1983.0637	187.21	0.1499	-0.1487	-0.0190
24	1983.7108	182.05	0.1456	-0.1455	-0.0054
25	1983.7135	179.56	0.1480	-0.1480	0.0010
26	1983.9337	181.00	0.1490	-0.1490	-0.0028
27	1983.9579	176.70	0.1570	-0.1568	0.0088
28	1984.0522	175.01	0.1446	-0.1441	0.0124
29	1984.0576	174.79	0.1445	-0.1439	0.0129
30	1984.0603	172.73	0.1355	-0.1344	0.0170
31	1984.7790	157.30	0.1460	-0.1348	0.0562
32	1984.9308	164.50	0.1410	-0.1359	0.0375
33	1985.1063	161.00	0.1370	-0.1296	0.0445
34	1985.2048	158.00	0.1250	-0.1160	0.0467
35	1985.8378	145.69	0.1141	-0.0943	0.0642
36	1985.8406	144.53	0.1202	-0.0980	0.0696
37	1985.8541	145.71	0.1200	-0.0992	0.0675

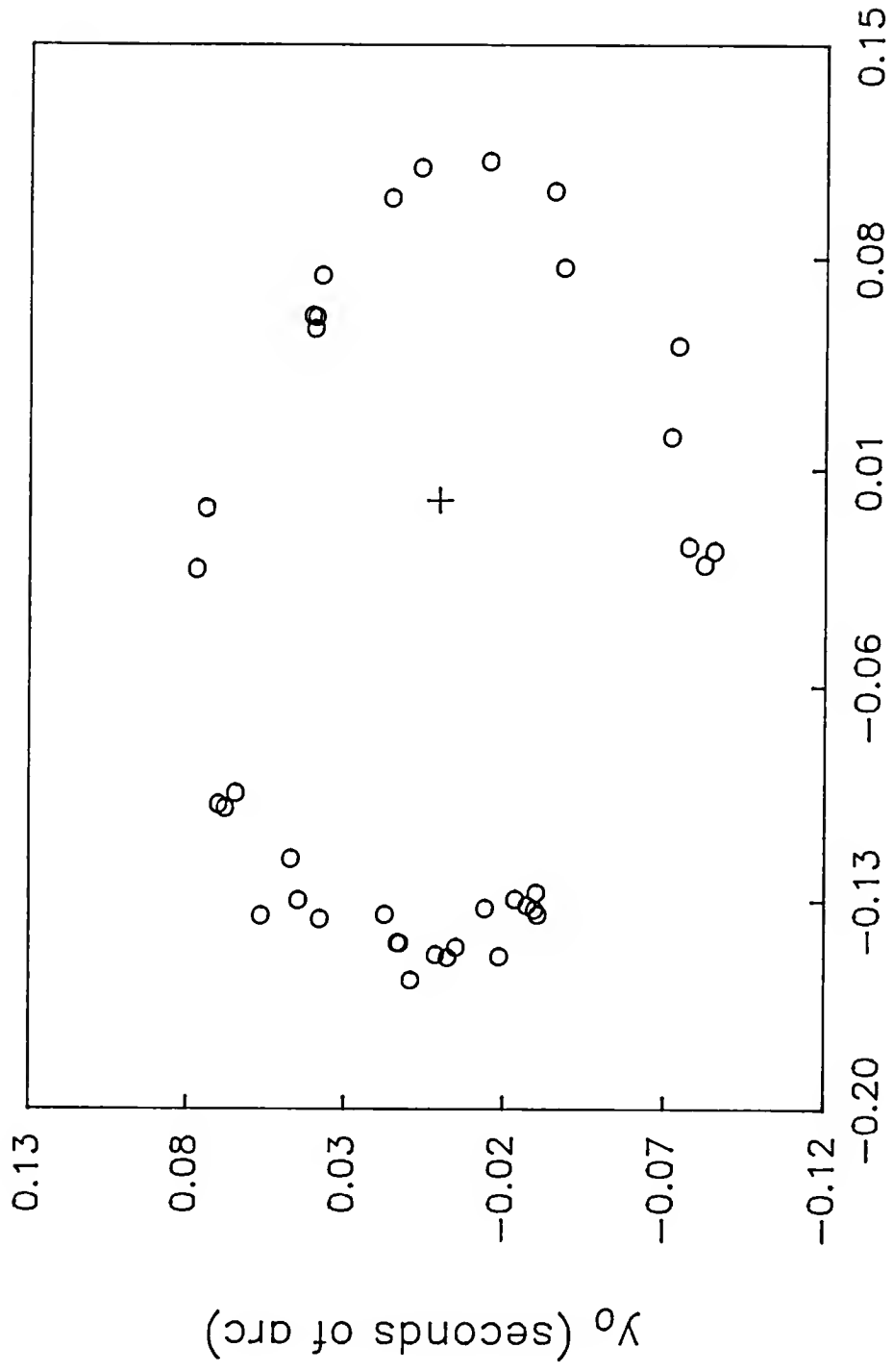

Figure 4-1. Plot of the observation data for 51 Tau in the x_0-y_0 plane.

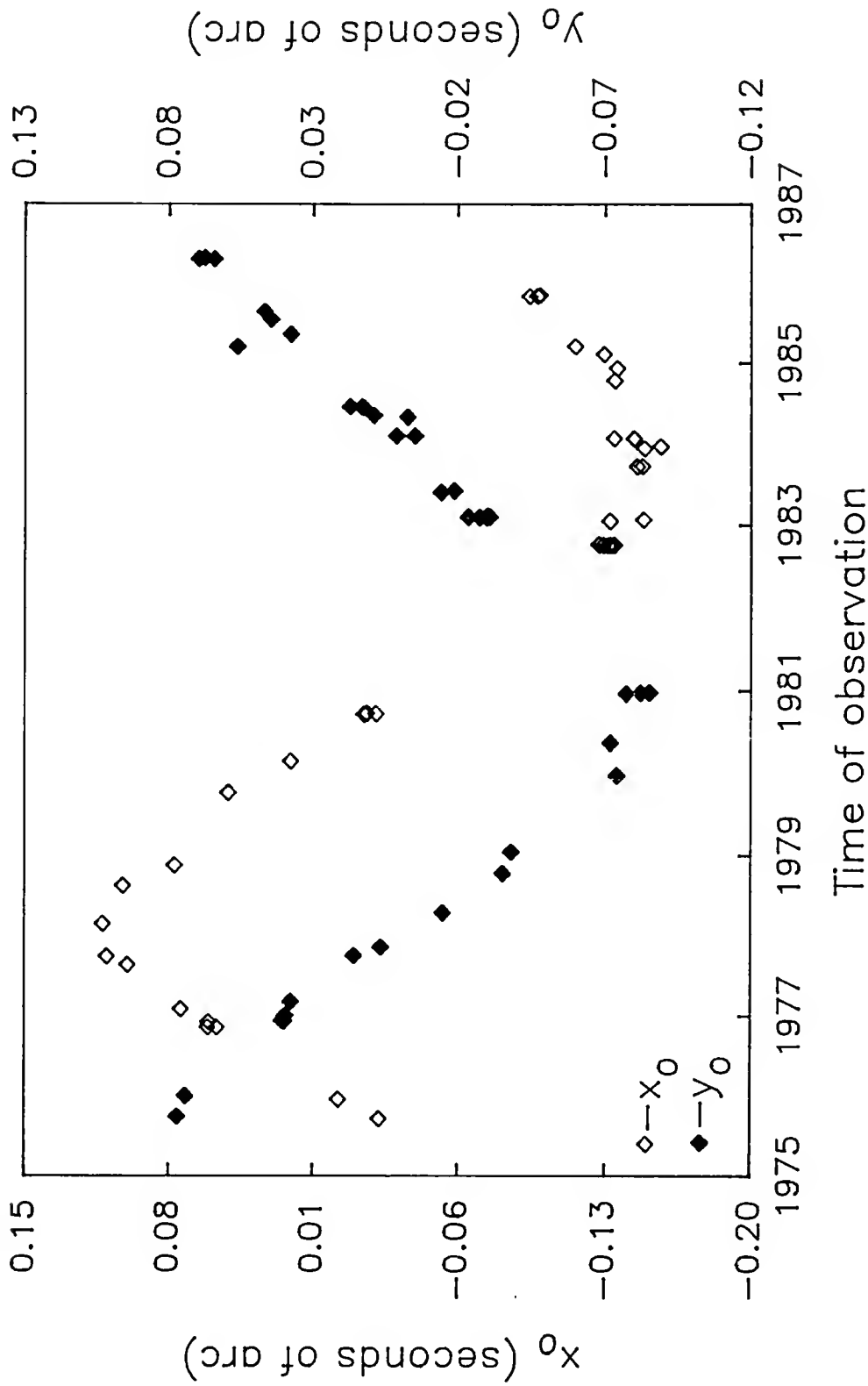

Figure 4-2. Plot of a) the x_0- b) y_0- coordinates against the observing epoch of the observation data for 51 Tau.

Table 4-5.

Initial Approximate Solution $\hat{a}_0$ for 51 Tau.

P_0(yr)	T_0	a_0''	e_0	$i_0°$	$\omega_0°$	$\Omega_0°$
11.18	1966.4	0.128	0.181	127.3	152.9	170.2

Table 4-6.

The Final Solution for 51 Tau

	P(yrs)	T	a"	e	i°	ω°	Ω°
the final solution	11.22	1966.5	0.128	0.173	125.5	157.3	171.2
standard deviations	0.039	0.031	0.0004	0.002	0.32	1.51	0.33
standard deviations of the uncorrelated parameters	0.00003	2.091	0.0002	0.0005	0.28	1.39	2.32
the efficiency				0.466			
the covariance matrix	.36E-06	.77E-09	.17E-06	.23E-06	-.12E-06	-.75E-06	-.66E-08
	.77E-09	.74E-11	.47E-09	-.12E-08	-.17E-09	.98E-09	.38E-09
	.17E-06	.47E-06	.30E-06	-.34E-06	-.14E-06	-.34E-06	-.66E-08
	.23E-06	-.12E-08	-.34E-06	.47E-05	.11E-06	-.14E-05	-.27E-06
	-.12E-06	-.17E-09	-.14E-05	.11E-06	.19E-06	.29E-06	-.16E-08
	-.75E-06	.98E-09	-.34E-06	-.14E-05	.29E-06	.29E-05	.30E-06
	-.66E-08	.38E-09	-.66E-08	-.27E-06	-.16E-08	.30E-06	.11E-06

Table 4-6 (continued)

	P(yrs)	T	a"	e	i°	ω°	Ω°
the correlation matrix	1.0000	0.4709	0.5131	0.1762	-0.4604	-0.7339	-0.0327
	0.4709	1.0000	0.3119	-0.2075	-0.1481	0.2099	0.4115
	0.5131	0.3119	1.0000	-0.2859	-0.6074	-0.3661	-0.0355
	0.1762	-0.2075	-0.2859	1.0000	0.1127	-0.3811	-0.3718
	-0.4604	-0.1481	-0.6074	0.1127	1.0000	0.3992	-0.0112
	-0.7339	0.2099	-0.3661	-0.3811	0.3992	1.0000	0.5273
	-0.0327	0.4115	-0.0355	-0.3718	-0.0112	0.5273	1.0000
the transformation matrix	-0.0074	0.9999	-0.0002	0.0000	0.0002	-0.0026	0.0033
	0.4926	0.0068	-0.2021	-0.0225	-0.1789	0.1955	-0.8036
	-0.1416	-0.0018	-0.5317	-0.0168	-0.8053	-0.0501	0.2145
	-0.4483	-0.0034	-0.6345	-0.1522	0.4443	-0.3079	-0.2848
	0.6780	0.0034	-0.4681	-0.0428	0.3145	0.0186	0.4691
	0.2546	0.0001	0.2331	-0.4862	-0.1520	-0.7869	-0.0466
	-0.1086	0.0003	0.0195	-0.8590	0.0121	0.4949	0.0703

Table 4-7.

Residuals of the Observations for 51 Tau in (ρ,θ) and (x,y)

	t	v_θ	v_ρ	v_x	v_y
1	1975.7160	1.0090	-0.0042	0.0000	-0.0044
2	1975.9591	1.1702	-0.0039	-0.0012	-0.0039
3	1976.8574	2.5044	0.0082	0.0048	0.0073
4	1976.8602	3.7620	0.0042	0.0007	0.0064
5	1976.9229	1.2477	0.0070	0.0050	0.0051
6	1977.0868	-0.0465	0.0009	0.0009	0.0002
7	1977.6398	-2.0624	-0.0015	-0.0010	-0.0039
8	1977.7420	0.5795	-0.0083	-0.0083	0.0004
9	1978.1290	0.2656	-0.0058	-0.0057	0.0011
10	1978.6183	-0.5497	-0.0015	-0.0018	-0.0006
11	1978.8756	-0.2196	0.0173	0.0152	-0.0082
12	1979.7735	-2.1815	-0.0054	-0.0059	0.0027
13	1980.1632	-1.2302	0.0036	-0.0008	-0.0039
14	1980.7182	-2.8268	-0.0005	-0.0038	0.0014
15	1980.7263	-0.0249	-0.0064	0.0014	0.0063
16	1980.7291	-3.4623	-0.0084	-0.0032	0.0093
17	1982.7550	1.9542	-0.0022	0.0031	-0.0038
18	1982.7579	1.0538	-0.0041	0.0045	-0.0013
19	1982.7605	3.2987	0.0007	0.0009	-0.0074
20	1982.7633	0.7102	-0.0058	0.0060	-0.0001
21	1982.7661	0.1717	0.0015	-0.0014	-0.0006
22	1983.0472	2.7391	0.0052	-0.0043	-0.0069
23	1983.0637	1.4488	-0.0111	0.0115	-0.0019
24	1983.7108	-2.8848	0.0019	-0.0020	0.0075
25	1983.7135	-0.4324	-0.0005	0.0005	0.0013
26	1983.9337	-4.9126	-0.0004	0.0007	0.0129
27	1983.9579	-0.9444	-0.0083	0.0085	0.0022
28	1984.0522	-0.5459	0.0042	-0.0040	0.0020
29	1984.0576	-0.3999	0.0043	-0.0042	0.0016
30	1984.0603	1.6232	0.0133	-0.0136	-0.0023
31	1984.7790	6.9626	-0.0028	-0.0031	-0.0173
32	1984.9308	-2.5201	-0.0004	0.0022	0.0060
33	1985.1063	-1.7784	0.0000	0.0015	0.0042
34	1985.2048	-0.3956	0.0097	-0.0086	0.0046
35	1985.8378	-0.2949	0.0015	-0.0008	0.0014
36	1985.8406	0.8016	-0.0047	0.0037	-0.0039
37	1985.8541	-0.6865	-0.0050	0.0050	-0.0016

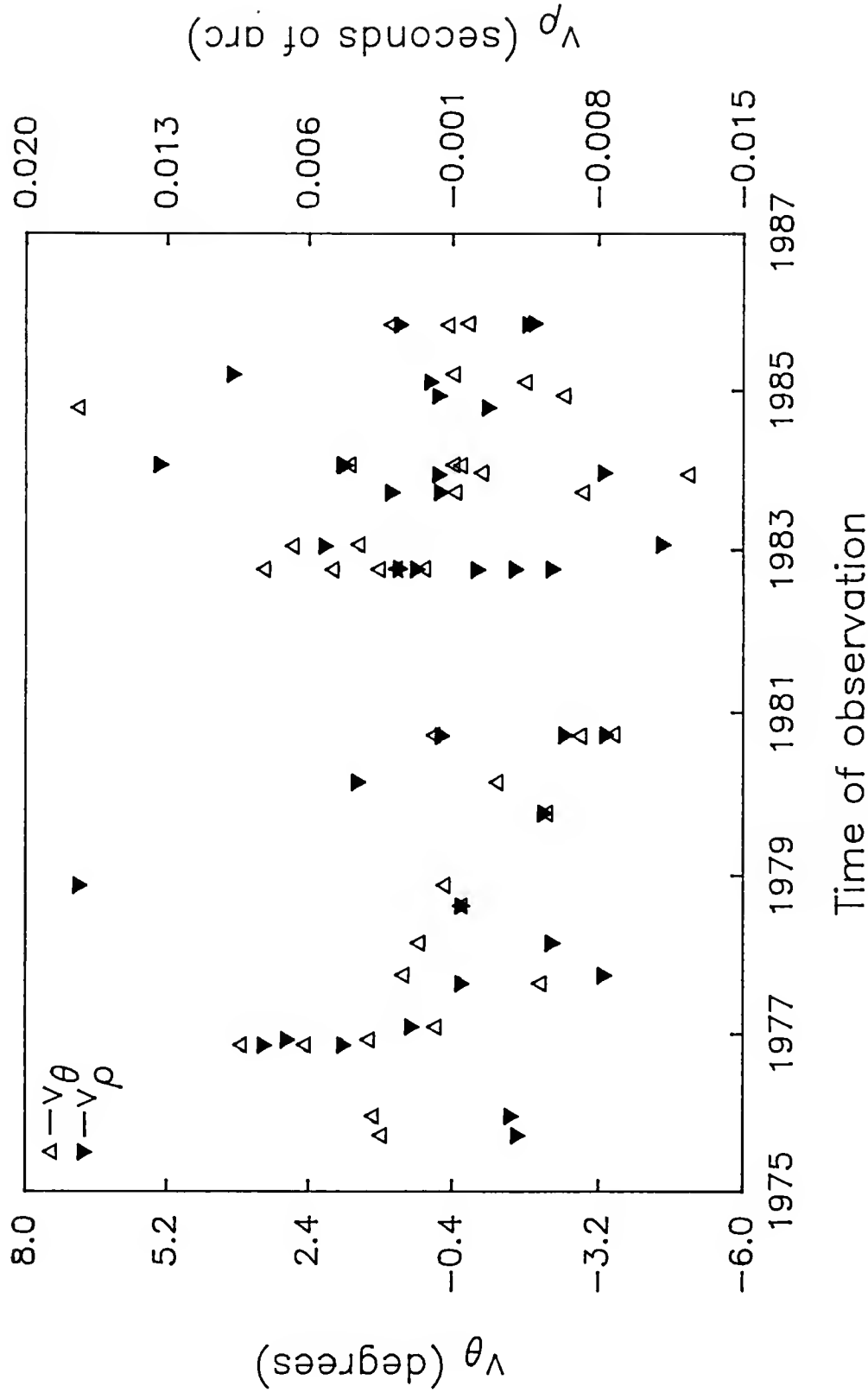

Figure 4-3. The residuals of the observations for 51 Tau in (ρ, θ).

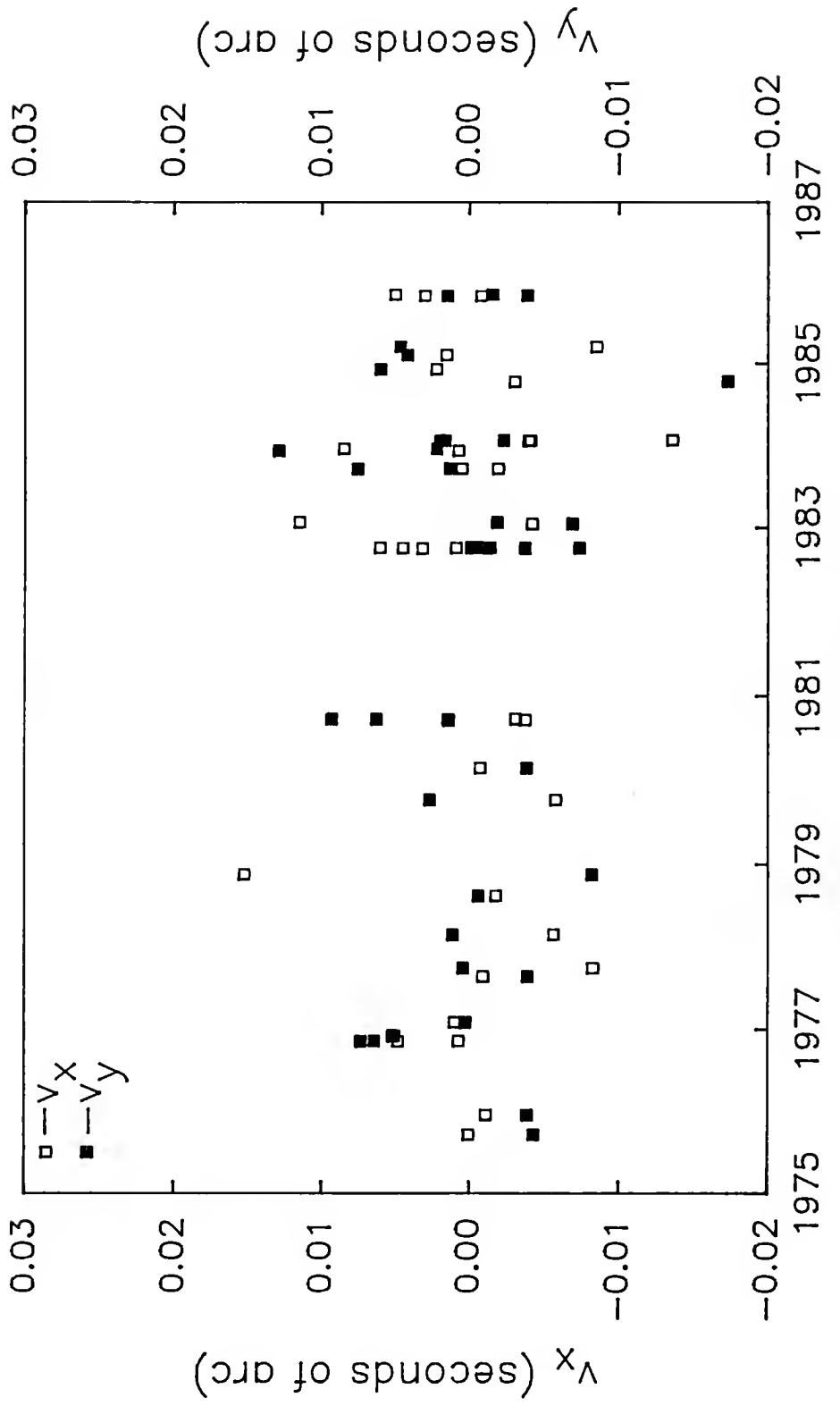

Figure 4-4. The residuals of the observations for 51 Tau in (x,y).

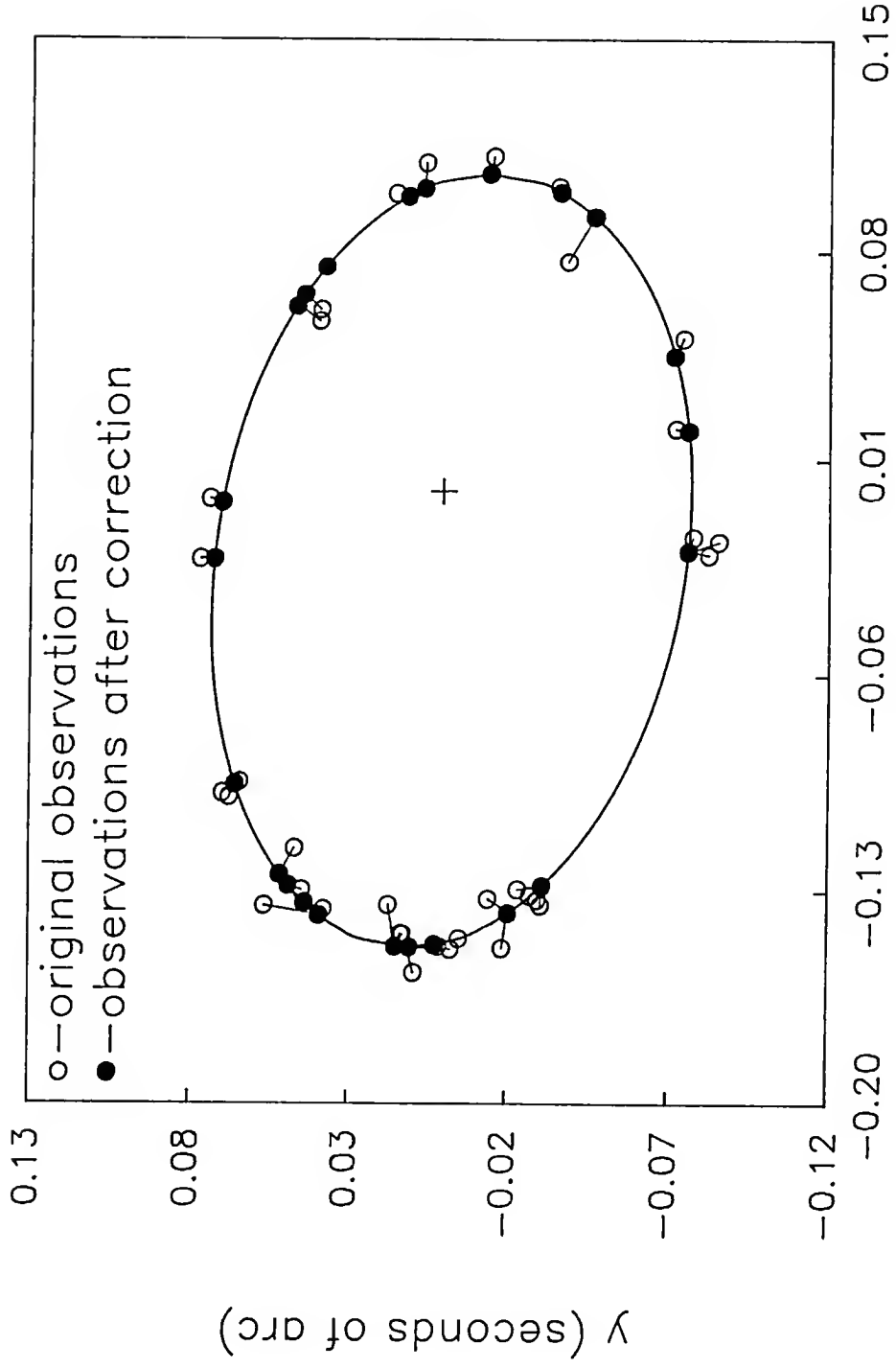

Figure 4-5. The original observations for 51 Tau compared with the observations after correction.

Remarks

Newton's method would converge well if the observational errors are small enough and the initial approximate solution is sufficiently accurate. But, when the errors in the observed distances and position angles are not sufficiently small, or the initial approximation is not close enough to the final solution, two things will happen:

1) In some iterations, the corrections to the adjustment parameters are too big so that some parameters go to unreasonable values; e.g., the semi-major axis a becomes negative or the mean motion n, the eccentricity e becomes negative, which are unacceptable; and further calculation becomes pointless.

2) The iterations do not converge; i.e., the residual function S becomes larger in the next iteration, although all parameters remain in the ranges of reasonable values.

In these cases, Newton's method will fail to yield a solution. In the next chapter, two other approaches (the modified Newton methods) are proposed to deal with these cases.

CHAPTER V
THE MODIFIED NEWTON SCHEME

The scheme for solving the nonlinear condition equations by Newton's method has been discussed in the last chapter. Although rapid convergence can be expected if the initial approximate solution is sufficiently accurate and the residuals are small enough, Newton's method often fails to yield a solution, particularly if the initial approximation $\hat{\mathbf{a}}_0$ of the vector $\hat{\mathbf{a}}$ is greatly in error or the residuals are very large. As mentioned in the last chapter, two problems arise in these cases:

1) Some of the corrections δ to the adjustment parameters are too big, which is unacceptable. For example, e_0=0.15, but δ_e=-0.16, so that the new value e=-0.01; in this case, the residual function calculated is no longer meaningful and any further calculation becomes pointless.

2) Divergence occurs directly; i.e., the value of residual function S_0 is larger after the iteration than it was before.

For both of these cases, the major problem is that the step size could be too large. If we decrease the step size, the situation might be improved to some extent. We still rely on Newton's method indicating the right direction, but

we no longer apply the full corrections which would follow from the original formalism. In this chapter, we will discuss the combination of Newton's method with the method of steepest descent and other sophisticated methods.

The Method of Steepest Descent

We retain $\mathbf{A}=\mathbf{f}_{\hat{a}}^{T}\mathbf{w}\mathbf{f}_{\hat{a}}$ and introduce a numerical factor $1/f$ into the normal equation (4-8), which thus becomes

$$\frac{1}{f}\mathbf{A}\delta = -\mathbf{f}_{\hat{a}}^{T}\mathbf{w}\hat{\phi} \quad . \tag{5-1}$$

By choosing an appropriate value of f, we would reduce the step size and make the residual function S_0 gradually smaller and smaller. This is analogous to the basic idea of "the method of steepest descent" which is to step to the next in a sequence of better approximations to the solution by moving in the direction of the negative of the gradient of S_0. If the step is not too large, the value of S_0 must necessarily decrease from one iteration to the next. Our goal is to arrive at the stable absolute minimum value of S_0 and to the corresponding set of parameters which is the best solution.

At this point, two new problems arise:

1) How does one calculate the residual function S_0 and its gradient?

2) How does one choose the best value of f?

Suppose we were to pretend that the current values of $\hat{\mathbf{a}}$ at any iteration are the true values of $\mathbf{a}$. This assumption is of course not true. But it makes it in general possible to estimate the vector $\hat{\mathbf{v}}$ as a function of $\hat{\mathbf{a}}$. This is so because if the present value of $\hat{\mathbf{a}}$ were the true value of $\mathbf{a}$, there would be a definite set of residuals $\hat{\mathbf{v}}$ which causes the corrected observations to satisfy the conditions equations rigorously. It is not difficult to find those components of $\hat{\mathbf{v}}$ which correspond to certain value of $\hat{\mathbf{a}}$. Suppose we have an approximation $\hat{\mathbf{a}}$ and $\mathbf{a}$ and $\hat{\mathbf{v}}_0$ of $\hat{\mathbf{v}}$, we wish to obtain the "best" approximation of $\hat{\mathbf{v}}$, assuming $\hat{\mathbf{a}}$ to be correct. We then need to minimize

$$S = \frac{1}{2}\hat{\mathbf{v}}^T\sigma^{-1}\hat{\mathbf{v}} + \mathbf{f}^T\hat{\boldsymbol{\mu}}$$

relative to the remaining variables $\hat{\mathbf{v}}$ and $\hat{\boldsymbol{\mu}}$ which yields

$$\mathbf{f}(\mathbf{x}_0 + \hat{\mathbf{v}}, \hat{\mathbf{a}}) = \mathbf{0} \qquad (5\text{-}2a)$$

$$\sigma^{-1}\hat{\mathbf{v}} + \mathbf{f}_{\hat{\mathbf{x}}}^T\hat{\boldsymbol{\mu}} = \mathbf{0} \qquad (5\text{-}2b)$$

Setting $\hat{\mathbf{v}}=\hat{\mathbf{v}}_0+\hat{\boldsymbol{\epsilon}}$, and expanding (5-2a) into powers of $\hat{\boldsymbol{\epsilon}}$, we have

$$\mathbf{f}(\hat{\mathbf{x}}_0, \hat{\mathbf{a}}) + \mathbf{f}_{\hat{\mathbf{x}}0}\hat{\boldsymbol{\epsilon}} = 0 \qquad (5\text{-}3)$$

and Eq. (5-2) becomes

$$\sigma^{-1}(\hat{\mathbf{v}}_0+\hat{\boldsymbol{\varepsilon}}) + \mathbf{f}_{\hat{\mathbf{x}}0}{}^{\mathrm{T}}\hat{\boldsymbol{\mu}} = 0 \qquad (5\text{-}4)$$

where $\hat{\mathbf{x}}_0 = \mathbf{x}_0 + \hat{\mathbf{v}}_0$. Solving Eq. (5-4) for $\hat{\boldsymbol{\varepsilon}}$ yields

$$\hat{\boldsymbol{\varepsilon}} = -\hat{\mathbf{v}}_0 - \sigma\mathbf{f}_{\hat{\mathbf{x}}0}{}^{\mathrm{T}}\hat{\boldsymbol{\mu}} \quad . \qquad (5\text{-}5)$$

Substituting now (5-5) into Eq. (5-3) for $\hat{\boldsymbol{\varepsilon}}$, we get

$$\mathbf{f}(\hat{\mathbf{x}}_0,\hat{\mathbf{a}}) - \mathbf{f}_{\hat{\mathbf{x}}0}\hat{\mathbf{v}}_0 - \mathbf{f}_{\hat{\mathbf{x}}0}\sigma\mathbf{f}_{\hat{\mathbf{x}}0}{}^{\mathrm{T}}\hat{\boldsymbol{\mu}} = 0. \qquad (5\text{-}6)$$

From Eq. (5-6) we obtain

$$\hat{\boldsymbol{\mu}} = \mathbf{w}[\mathbf{f}(\hat{\mathbf{x}}_0,\hat{\mathbf{a}}) - \mathbf{f}_{\hat{\mathbf{x}}0}\hat{\mathbf{v}}_0] \quad . \qquad (5\text{-}7)$$

From Eqs. (5-5) and (5-7) we arrive therefore at the expression

$$\hat{\mathbf{v}} = -\sigma\mathbf{f}_{\hat{\mathbf{x}}0}\mathbf{w}[\mathbf{f}(\hat{\mathbf{x}}_0,\hat{\mathbf{a}}) - \mathbf{f}_{\hat{\mathbf{x}}0}\hat{\mathbf{v}}_0] \quad , \qquad (5\text{-}8)$$

where $\mathbf{w}=\mathbf{f}_{\hat{\mathbf{x}}0}{}^{\mathrm{T}}\sigma\mathbf{f}_{\hat{\mathbf{x}}0}$ and $\hat{\mathbf{x}}_0 = \mathbf{x}_0 + \hat{\mathbf{v}}_0$ still. This equation may be iterated, if needed, to arrive at a definite $\hat{\mathbf{v}}$; i.e., until the value of $\hat{\mathbf{v}}$ on the left hand side and the value of $\hat{\mathbf{v}}_0$ on the right hand side are the same. There is thus a

corresponding $\hat{\mathbf{v}}$ to each choice of parameters $\hat{\mathbf{a}}$ and we can write $\hat{\mathbf{v}}=\hat{\mathbf{v}}(\hat{\mathbf{a}})$. Although Eq. (5-8) formally depends on $\hat{\mathbf{v}}_0$ in the first order, it is in fact rather insensitive to the actual value of $\hat{\mathbf{v}}_0$ used, since it is really quadratic in $\hat{\mathbf{v}}_0$. To appreciate this, note that

$$\mathbf{f}(\hat{\mathbf{x}}_0,\hat{\mathbf{a}}) - \mathbf{f}_{\hat{\mathbf{x}}0}(\hat{\mathbf{x}}_0-\mathbf{x}_0) = \mathbf{f}(\mathbf{x}_0,\hat{\mathbf{a}}) + \mathbf{O}(\hat{\mathbf{v}}_0{}^2) \quad . \qquad (5\text{-}9)$$

After we get the best value of $\hat{\mathbf{v}}$ which corresponds to a definite value of $\hat{\mathbf{a}}$, S is given by

$$S = \frac{1}{2}\,\hat{\mathbf{v}}^T\sigma^{-1}\hat{\mathbf{v}} + \hat{\mathbf{f}}^T\hat{\mu} = \frac{1}{2}\,\hat{\mathbf{v}}^T\sigma^{-1}\hat{\mathbf{v}} \quad ;$$

i.e., $S = S_0$,

because in this case, we already have $\hat{\mathbf{f}}=0$ for this value of $\hat{\mathbf{v}}$. Fixing the value of $\hat{\mathbf{a}}$, we can therefore get the best corresponding value of $\hat{\mathbf{v}}$, then calculate the value of the residual function S_0 which corresponds to this value of $\hat{\mathbf{a}}$. In other words, we may also write $S_0=S_0(\hat{\mathbf{a}})$.

Furthermore, when we calculate the best value of $\hat{\mathbf{v}}$ at $\hat{\mathbf{a}}$ from Eq. (5-8),

$$\hat{\mathbf{v}} = -\,\sigma\mathbf{f}_{\hat{\mathbf{x}}0}\mathbf{w}\hat{\phi} \quad ,$$

where again $\hat{\phi} = \mathbf{f}(\hat{\mathbf{x}},\hat{\mathbf{a}})-\mathbf{f}_{\hat{\mathbf{x}}}\hat{\mathbf{v}}$.

Hence, the expression for S_0 will be

$$S_0 = \frac{1}{2}\hat{\mathbf{v}}^T\sigma^{-1}\hat{\mathbf{v}} = \frac{1}{2}\hat{\boldsymbol{\phi}}^T\mathbf{w}\mathbf{f}_{\hat{\mathbf{x}}}\sigma\sigma^{-1}\sigma\mathbf{f}_{\hat{\mathbf{x}}}^T\mathbf{w}\hat{\boldsymbol{\phi}} = \frac{1}{2}\hat{\boldsymbol{\phi}}^T\mathbf{w}\hat{\boldsymbol{\phi}} \quad . \qquad (5\text{-}10)$$

This expression is exact if it is evaluated with the optimum value of $\hat{\mathbf{v}}$. In fact, however,

$$\hat{\boldsymbol{\phi}} = \mathbf{f}(\hat{\mathbf{x}},\hat{\mathbf{a}}) - \mathbf{f}_{\hat{\mathbf{x}}}\hat{\mathbf{v}} \approx \mathbf{f}(\mathbf{x}_0,\hat{\mathbf{a}})$$

is not sensitive to the value of $\hat{\mathbf{v}}$ used.

From Eq. (5-10), we can estimate the negative gradient δ_g of S_0 with respect to $\hat{\mathbf{a}}$:

$$-\delta_g = \frac{\partial S_0}{\partial \hat{\mathbf{a}}} = \frac{\partial \hat{\boldsymbol{\phi}}^T}{\partial \hat{\mathbf{a}}}\mathbf{w}\hat{\boldsymbol{\phi}} + \frac{1}{2}\hat{\boldsymbol{\phi}}^T\frac{\partial \mathbf{w}}{\partial \hat{\mathbf{a}}}\hat{\boldsymbol{\phi}}$$

$$= \left(\mathbf{f}_{\hat{\mathbf{x}}}\frac{\partial \hat{\mathbf{v}}^T}{\partial \hat{\mathbf{a}}} + \mathbf{f}_{\hat{\mathbf{a}}}^T - \mathbf{f}_{\hat{\mathbf{x}}}^T\frac{\partial \hat{\mathbf{v}}^T}{\partial \hat{\mathbf{a}}} - O(\hat{\mathbf{v}}) \right)\mathbf{w}\hat{\boldsymbol{\phi}} + O(\hat{\mathbf{v}}^2)$$

$$\approx \mathbf{f}_{\hat{\mathbf{a}}}^T\mathbf{w}\hat{\boldsymbol{\phi}} \quad , \qquad (5\text{-}11)$$

an expression which also depends only weakly on $\hat{\mathbf{v}}$. In deriving Eq. (5-11), the symmetry of $\mathbf{w}$ has been used to combine terms, and terms $O(\hat{\mathbf{v}}^2)$ have been neglected. The vector δ_0 points along the negative gradient of S_0, i.e., the direction of steepest descent. At this point, let us

delineate the method of steepest descent in our problem first. It can be modified as follows.

At each step, use the current values of $\hat{\mathbf{v}}$ and $\hat{\mathbf{a}}$ to compute

$$\hat{\boldsymbol{\phi}} = \mathbf{f}(\hat{\mathbf{x}},\hat{\mathbf{a}}) - \mathbf{f}_{\hat{\mathbf{x}}}\hat{\mathbf{v}} \quad , \tag{5-12a}$$

$$\hat{\mathbf{v}} = - \sigma \mathbf{f}_{\hat{\mathbf{x}}}^{\mathrm{T}}\hat{\boldsymbol{\phi}} \quad , \tag{5-12b}$$

and iterate Eqs. (5-12) above to get the best value $\hat{\mathbf{v}}$ at $\hat{\mathbf{a}}$. Since $\hat{\boldsymbol{\phi}}$ is quite insensitive to the values of $\hat{\mathbf{v}}$, this should converge rapidly. Then calculate the residual function at the best value of $\hat{\mathbf{v}}$, using

$$S_0 = \frac{1}{2}\,\hat{\mathbf{v}}^{\mathrm{T}}\sigma^{-1}\hat{\mathbf{v}} \qquad \text{or} \quad \frac{1}{2}\,\hat{\boldsymbol{\phi}}^{\mathrm{T}}\mathbf{w}\hat{\boldsymbol{\phi}} \quad . \tag{5-12c}$$

The corrections to the parameters are next computed from

$$\delta_g = - f\mathbf{f}_{\hat{\mathbf{a}}}^{\mathrm{T}}\mathbf{w}\hat{\boldsymbol{\phi}} \quad , \tag{5-12d}$$

$$\hat{\mathbf{a}}_n = \hat{\mathbf{a}} + \delta_g \quad . \tag{5-12e}$$

Now $\hat{\mathbf{a}}_n$ is available. Using the procedure (5-12a) to (5-12c) again, we can get the new value S_0. In Eq. (5-12d), the proportionality constant f must be chosen judiciously so that all components of $\hat{\mathbf{a}}_n$ do not exceed reasonable ranges and the new value of S_o is smaller than the old one.

This procedure can always be made to converge. But, as we know, the method of steepest descent too has its drawbacks, principally because typically the convergence may be rather slow.

The Combination of Newton's Method with the Method of Steepest Descent--The Modified Newton Scheme

Because of the drawbacks of the method of steepest descent we would like to incorporate the results above into alternative algorithms which combine the best of the Newton's method with the method of steepest descent by "interpolating" between them; i.e, combine the certain improvement obtained by the steepest descent method when the available approximation is far from the definitive solution, with the rapid convergence of Newton's method when the current approximation is already close to it.

In order to implement this idea, Eq. (5-12d) must be slightly modified.

Consider the equation

$$\frac{1}{f} D\delta = - f_{\hat{a}}^T w\hat{\phi} \tag{5-13}$$

and compare it with Eqs. (5-12d) and (4-8). In Eq. (5-12d), $D=I$, a unit matrix, and in Eq. (4-8), $f=1$ and $D=A=f_{\hat{a}}^T w\hat{\phi}$. The equation (5-12d) leads to the pure steepest descent solution, which turns out to be undesirable in general. Actually, we can retain the flexibility to choose D

more freely. As Jefferys remarked, this will allow us to improve certain characteristics of the solution.

Set $\mathbf{D=A=f_{\hat{a}}^T w\hat{\phi}}$ in Eq. (5-13) to obtain

$$\frac{1}{f}\mathbf{A\delta} = -\mathbf{f_{\hat{a}}^T w\hat{\phi}}$$

which is the same as Eq. (5-1).

If Eq. (5-1) is used instead of (5-12d), the only difference between the normal equations of Newton's method and Eq. (5-1) is the numerical factor f.

We adopt a modification of Newton's method which introduces only the numerical factor f into the normal equation (4-8) as in Eq. (5-1). In this algorithm, we choose the value of f such that $S_0(=\hat{\phi}^T w\hat{\phi})$, evaluated as a function of f, is a minimum. Once δ has been chosen in a current iteration, that value of f is determined which minimizes--for this iteration--the value of S_0. It is plausible that this optimizes the gain of accuracy during this iteration, and the next iteration proceeds from there. In other words, we use the usual normal equations (f=1) to estimate the direction of the vector at each step, and then choose the value of f to get the optimum length.

For arriving at the optimum value of f, the so-called "Golden section method" or "0.618 method" is used. This method was originally conceived for searching for the minimum of a function of one variable. The absolute minimum

may be found by this method if the function has only a single peak or a strict vertex. Otherwise, for a function with more than one peaks, the method will find a local minimum. In any case, a point at which the function is smaller than at the initial point can be always reached.

For every set of parameters $\hat{\mathbf{a}}$, there is a corresponding best set of corrections $\hat{\mathbf{v}}$, and for every value of f, there is a corresponding set of corrections $\boldsymbol{\delta}$. Therefore, actually we can write $S_0=S_0(f)$; the scalar residual function S_0 is a function of variable f. Anyway, we can thus obviously search at least for that f which will produce a local minimum of S_0.

The scheme for finding best value of f (i.e. the minimum S_0) at each iteration consists of two parts. The first of these is to delineate an interval in which the minimum is located; then look for the minimum by the "0.618 method" within this interval. The searching procedure is as follows.

1) Choose u as the length for the initial step as well as a positive number $F=(\sqrt{5}-1)/2$ ($F\approx 0.618$) and a small quantity ε.

2) Define $f=10^u$; $S_0=S_0(f)$ now is equivalent to $S_0=S_0(u)$; put $u^{(0)}=0$. The case u=0, i.e., $f=1$, corresponds to Newton's method. The superscript indicates the sequence number of the current iteration.

If $S_0(u^{(0)} + u)\leq S_0(u^{(0)})$, go to step 3.

If $S_0(u^{(0)} + u)>S_0(u^{(0)})$, put

$u = -u$,

$u^{(-1)} = u^{(0)} + u$,

and also go to step 3.

3) Calculate $u^{(k+1)}=u^{(k)} + u$ and $S_0(u^{(k+1)})$.

4) If $S_0(u^{(k+1)}) \le S(u^{(k)})$, then

$u=2u$; $K=K+1$; go to step 3.

If $S_0(u^{(k+1)}) > S(u^{(k)})$, then

$u_3^{(0)} = u^{(k+1)}$, $u_1^{(0)} = u^{(k-1)}$;

$d^{(0)} = u_3^{(0)} - u_1^{(0)}$;

$i = 0$;

go to step 5.

5) Let $y_1^{(i)} = u_3^{(i)} - Fd^{(i)}$, $y2^{(i)} = u_1^{(i)} + Fd^{(i)}$.

6) If $S_0(y_1^{(i)}) < S_0(y_2^{(i)})$, then

$d^{(i+1)} = y_2^{(i)} - u_1^{(i)}$, $u_1^{(i+1)} = u_1^{(i)}$,

$u_3^{(i+1)} = y_2^{(i)}$ and

go to step 7.

If $S_0(y_1^{(1)}) > S_0(y_2^{(1)})$, then

$d^{(i+1)} = u_3^{(i)} - y_1^{(i)}$, $u_1^{(i+1)} = y_1^{(i)}$,

$u_3^{(i+1)} = u_3^{(i)}$ and

also go to step 7.

If $S_0(y_1^{(i)} = S_0(y_2^{(i)})$, then

$d^{(i+1)} = y_2^{(i)} - u_1^{(i)} = u_3^{(i)} - y_1^{(i)}$ and

$u_1^{(i+1)} = u_1^{(i)}$, $u_3^{(i+1)} = y_2^{(i)}$ or

$u_1^{(i+1)} = y_1^{(i)}$, $u_3^{(i+1)} = u_3^{(i)}$

and also go to step 7.

7) If $|d^{(i+1)}| \le \varepsilon$, then stop and

$u_{min} = u_1^{(i+1)} + d^{(i+1)}/2$;

Otherwise i=i+1 and return to step 5.

This process searches for the minimum of S_0 at each iteration and accelerates the convergence.

In summation, Newton's modified method consists of the following steps.

Step 1.

Using Eqs. (5-12a) and (5-12b) iteratively, get the best value of $\hat{\mathbf{v}}$ at $\hat{\mathbf{a}}$ (initially, set $\hat{\mathbf{a}}=\hat{\mathbf{a}}_0$ and since $\hat{\mathbf{v}}$ is unknown, set $\hat{\mathbf{v}}_0=0$). This converges rather rapidly.

Step 2.

Compute $S_0=S_0(\hat{\mathbf{a}})=\hat{\boldsymbol{\phi}}^T\mathbf{w}\hat{\boldsymbol{\phi}}$ at $\hat{\mathbf{a}}$.

Step 3.

Set the starting value for f equal to 1, i.e., u=0 and use Eq. (5-1) (in this case, equivalent to Eq. (4-8)), to compute $\boldsymbol{\delta}$.

Step 4.

Calculate $\hat{\mathbf{a}}_n$ from $\hat{\mathbf{a}}_n = \hat{\mathbf{a}} + \boldsymbol{\delta}$.

Step 5.

Check every element in $\hat{\mathbf{a}}_n$ to see if they are all located in the ranges of the allowed values (e.g., a>0, 0<e<1, n>0,...). If not, reduce u=u-0.5 (i.e., $f=10^{u-0.5}$) and recompute the correction vector $\boldsymbol{\delta}$ using the equation

$$\boldsymbol{\delta}' = f\boldsymbol{\delta} \quad , \tag{5-14}$$

and set $\delta=\delta'$, return to step 4.

This step may be iterated until every element in $\hat{a}$ has converged to a reasonable value.

Step 6.

Repeat the process of step 1 to get the best $\hat{v}$ at $\hat{a}_n$ and compute

$S_0' = S_0(\hat{a}_n) = S_0(\hat{a}+\delta)$.

If $S_0' \geq S_0$, set u=u-0.5 and return to step 5.

Step 7.

If $S_0' < S_0$ and u=0 (f=1), then go to step 9.

If $S_0' < S_0$ and u≠0 (f≠1), then proceed to next step.

Step 8.

Use the "0.618 method" to find the optimum value of f to get the minimum of the residual function, S_{min}, at this iteration, and the corresponding optimum improvement of $\hat{a}_n$.

Step 9.

Test for convergence. That is, test again the size of each component of δ and $\hat{v}$ against the corresponding component in $\hat{a}$ and $\hat{x}$. When the change is sufficiently small for all components, the process may be said to have converged. If the convergence has not yet been achieved, return to step 3 for next iteration.

The scheme described above is efficient. In comparison with Newton's method which was described in the last chapter, we find two essential differences:

1) In this method, a corresponding best estimate $\hat{\mathbf{v}}$ has to be found for every estimate $\hat{\mathbf{a}}$ by using Eq. (5-12a) and (5-12b) iteratively. In Newton's method, this is not taken into account.

2) In the method proposed here, the numerical factor f is applied to the normal equation for δ and the best value of f, which leads to a minimum S_{min}, must be found by the "0.618 method".

We see, however, that if the $\hat{a}_n$, which is computed from that value of δ which is the solution of the same normal equations for δ as in Newton's method (here, f=1), already produces a smaller value of S_0, the scaler residual function, one does not need to bother with searching for the best value of f for the iteration. This iteration is then similar to what one would do in Newton's method.

Jefferys (1981) points out that this method is somewhat like a modified Fletcher-Powell algorithm (1963). We therefore call this as "FP method". In addition, as a slight further modification to scheme above, Marquardt's algorithm (D. W. Marquardt, 1963) can also be incorporated to suggest another approach for solving our orbit problem.

The Application of Marquardt's Algorithm

In a sense, the basic idea behind Marquardt's algorithm is to interpolate between Newton's method and the method of steepest descent such that initially, far from the solution, steepest descent dominates, and that the more efficient

Newton method dominates the calculations as the solution is approached. This is similar to what we have done above in our improved Newton's method.

Comparison of Marquardt's paper (1963) with Eq. (5-1) suggests that the natural generalization of this fundamental equation is

$$(\mathbf{A} + \frac{1}{f}\mathbf{D})\,\boldsymbol{\delta} = -\,\mathbf{f}_{\hat{a}}^{T}\mathbf{w}\hat{\boldsymbol{\phi}} \quad , \tag{5-15}$$

where **D** is to be chosen appropriately. Because the gradient methods are not scale invariant, it is desirable to choose **D** so that it has the form

$$\mathbf{D} = \text{diag}\,(A_{11}, A_{22}, \ldots, A_{77}), \tag{5-16}$$

where A_{jj} are the elements of the matrix **A**. The effect of this is to scale the gradient along each coordinate axis by the factor $A_{jj}^{-1/2}$.

This is also exactly Jefferys' (1981) suggestion. In our case, Marquardt's algorithm would take the same form as described above for the improved Newton scheme, namely the "FP method", except for step 3, where Eq. (5-1) would have to be replaced by

$$(\mathbf{A} + \frac{1}{f}\mathbf{D})\boldsymbol{\delta} = -\mathbf{f}_{\hat{a}}^{T}\mathbf{w}\hat{\boldsymbol{\phi}} \quad .$$

We call this the "MQ method".

Using the two modifications of Newton's method just described (FP and MQ method, respectively) for all observation data, the same result should be obtained. Our calculation confirmed this for all the practical examples in our paper. We noticed, however, that Marquardt's method converges much faster than the other one. The reason for this might be as follows.

Comparing the two equations (5-1) and (5-15):

$$\frac{1}{f}\mathbf{A}\boldsymbol{\delta} = -\mathbf{f}_{\hat{a}}^{T}\mathbf{w}\hat{\boldsymbol{\phi}} \qquad \text{and}$$

$$(\mathbf{A} + \frac{1}{f}\mathbf{D})\boldsymbol{\delta} = -\mathbf{f}_{\hat{a}}^{T}\mathbf{w}\hat{\boldsymbol{\phi}} \quad .$$

where $\mathbf{A}=\mathbf{f}_{\hat{a}}^{T}\mathbf{w}\mathbf{f}_{\hat{a}}$ and $\mathbf{D}=\mathrm{diag}(A_{11}, A_{22}, \ldots, A_{77})$, we see that in equation (5-1), which is used in "FP Method", the only difference from $\mathbf{A}\boldsymbol{\delta}=-\mathbf{f}_{\hat{a}}^{T}\mathbf{w}\hat{\boldsymbol{\phi}}$, the normal equation in the Newton scheme, is the numerical factor f; all elements in $\mathbf{A}$ are multiplied by the same factor $1/f$. The solution $\boldsymbol{\delta}$ is therefore only changed by the same numerical factor, i.e., the ratios between all elements in $\boldsymbol{\delta}$ remain unchanged. However, in the "MQ method", as seen in Eq. (5-15), the coefficient matrix of $\boldsymbol{\delta}$ is modified as

$$\mathbf{A} + \frac{1}{f}\mathbf{D} \quad ,$$

in which only the diagonal elements are changed. The coefficient matrix of δ is gradually becoming diagonal in the course of the "MQ" iterations. This means that the corrections to the parameters (but not the parameters themselves) become uncorrelated. However, we have to notice that the inverse of the coefficient matrix of δ and the covariance matrix of the parameters are no longer the same in this method, since the covariance matrix of the parameters is intrinsically the inverse of **A**. Thus we calculate only the "efficiency" of the adjusted parameters in final solution.

Two Practical Examples

Table 5-1 lists the observations for β738. These observations were collected by Heintz. Table 5-2 contains the initial data. All the position angles in the observations have been reduced to equator 2000.0. Figure 5-1 shows all the observation points plotted in the x_0-y_0 plane. In Figure 5-2, the x_0, y_0 coordinates are respectively plotted against the observing epoch. From Figure 5-2, we see that the four earliest observations are scattered widely. The initial approximate solution is listed in Table 5-3.

Initially, the same weights were assigned to all 26 observations. The same final solution is obtained using both the "FP" and the "MQ" scheme as shown in Table 5-4; this is called "solution #1" for β738. "FP" requires 13 iterations and spent $12^{m}06^{s}.18$ CPU time on the VAX while "MQ"

requires only 10 iterations and $9^{m}11^{s}.01$ CPU time for arriving at the same final result. The covariance, the correlation and the transformation matrices as well as the standard deviations of the parameters in this solution are also listed in Table 5-4.

For evaluating a final result, the residuals are important. They must be random and reasonably small. The residuals in ρ, θ, x and y for this solution are listed in Table 5-5, which shows that some are very big, especially those for the four earliest points which were observed before 1921. In Figures 5-3 and 5-4, the residuals in ρ, θ, x and y are respectively plotted against the observing epoch. Also, Figure 5-5 displays all the initial observation points compared with those after correction; short bars connect every two corresponding points. From the relevant tables and plots above, we see that the residuals seem to be not very reasonable.

Heintz had already computed an orbit for β738 from the same observation data. His result is shown in Table 5-6 which differs greatly from the solution #1. This discrepancy originates in earliest four uncertain observations. These weak observations affect the overall result. Giving all the observations the same weights is obviously inappropriate. We therefore assigned a much lower weight (0.03) to the earliest four observations. Using these weighted observation data, both the "FP" and "MQ" scheme

still yield exactly the same solution which agrees fairly well with Heintz' result. This solution is listed in Table 5-7 and called "solution #2" for β738. This time, both methods run through only 6 iterations and used the same CPU time, around 2^m on the VAX, because in all iterations by both methods Newton's method dominates, that is, neither "FP" nor "MQ" has been really called into action. The residuals in ρ, θ, x and y in this solution are generally much smaller than in solution #1, which are listed in Table 5-8, and plotted in Figures 5-6 and 5-7. Figure 5-8 shows again the comparison of initial observations with those after correction. Table 5-7 shows also the final covariance, correlation and transformation matrices as well as the standard deviations.

In addition, we also removed the four observations before 1921 from the input data. Using only the 22 observations that remained, the solution is essentially the same as the weighted 26 observations by both the methods as we would expect.

Another example is BD+19°5116. The observations (35 points) and the reduced initial data are shown in Tables 5-9 and 5-10. The observations are again plotted in the x_0-y_0 plane in Figure 5-9, and x_0, y_0 coordinates are also plotted against the time of observation in Figure 5-10. From Figure 5-7, we see that the observations cover only a very short

arc on the ellipse. The computation from these data will be more difficult.

Using the "MQ" method, we arrived at a final solution which is shown in Table 5-12. Only 5 iterations are needed and around 9^m CPU time is used. The residuals in ρ, θ, x and y are listed in Table 5-13, and plotted in Figures 5-11 and 5-12. The initial observations are again compared with those after correction in Figure 5-13. Also, the covariance, the correlation and the transformation matrices and the standard deviations are given in Table 5-12.

For these data, the "FP" scheme was also used. Although it gave the same result, it ran through around 900 iterations and required 4 hours CPU time!

Table 5-1.

The Observation Data for β738
(courtesy W. D. Heintz)

β 738	02232	s2952	(2000)	7.5-7.8 F8
1879.70	183.1	0.64	2	β
1891.80	174.7	0.55	3	β
1900.21	177.5	.75	3	Doo 2 Booth 1
05.98	193.8	.66	2	01
21.15	50.5	.44	10	δ
26.67	43.5	.46	15	V 8 B 7
29.77	40.0	.50	8	B V 4
32.45	38.5	.55	15	φ 8 B 4 δ 3
34.76	37.3	.57	11	B V 4 δ 3
38.16	35.0	.59	14	Sim 10 B 4
42.52	32.3	.57	7	B 4 V 3
45.80	30.1	.62	12	Strom 8 V 4
47.98	30.4	.44	8	B
55.00	25.6	.22	4	B 2 φ int 2
56.45	13.2	.19	4	B 2 φ int 2
59.03	352.1	.13	10	φ int 7 B 3
59.97	307.1	.10	6	φ int 3 B 3
61.03	249.2	.11	3	φ int
62.53	222.7	.14	2	φ int
64.58	223.5	.19	14	φ int 10 B 4
66.02	217.5	.26	4	φ int 2 B 2
66.94	220.9	.32	5	Kni
69.04	218.8	.56	3	φ int
75.55	217.4	.62	9	Wor 7 Hln 2
77.28	217.1	0.75	7	Hln 4 hz 3
1983.71	215.4	1.03	2	hz

Finsen published an elliptical (P=110yr) and a parabolic solution (cf the 1969 orbit catalogue). Heintz' result is: P 290 yr, T 1952.0, a 1.460, e 0.63, i 96.7, ω 38.0, node 29.4 (2000).

The position angles are oriented to equator 2000. The observations before 1920 are from Northern instruments, and are correspondingly uncertain.

Table 5-2.

The Reduced Initial Data for β738

	t	θ_0	ρ_0	x_0	y_0
1	1879.70	183.1	0.64	-0.639	-0.035
2	1891.80	174.7	0.55	-0.548	0.051
3	1900.21	177.5	0.75	-0.749	0.033
4	1905.98	193.8	0.66	-0.641	-0.157
5	1921.25	50.5	0.44	0.280	0.340
6	1926.67	43.5	0.46	0.334	0.317
7	1929.77	40.0	0.50	0.383	0.321
8	1932.45	38.5	0.55	0.430	0.342
9	1934.76	37.3	0.57	0.453	0.345
10	1938.16	35.0	0.59	0.483	0.338
11	1942.52	32.3	0.57	0.482	0.305
12	1945.80	30.1	0.62	0.536	0.311
13	1947.98	30.4	0.44	0.380	0.223
14	1955.00	25.6	0.22	0.198	0.095
15	1956.45	13.2	0.19	0.185	0.043
16	1959.03	352.1	0.13	0.129	-0.018
17	1959.97	307.1	0.10	0.060	-0.080
18	1961.03	249.2	0.11	-0.039	-0.103
19	1962.53	222.7	0.14	-0.103	-0.095
20	1964.58	223.5	0.19	-0.138	-0.131
21	1966.02	217.5	0.26	-0.206	-0.158
22	1966.94	220.9	0.32	-0.242	-0.210
23	1969.04	218.8	0.56	-0.436	-0.351
24	1975.55	217.4	0.62	-0.493	-0.377
25	1977.28	217.1	0.75	-0.598	-0.452
26	1983.71	215.4	1.03	-0.840	-0.597

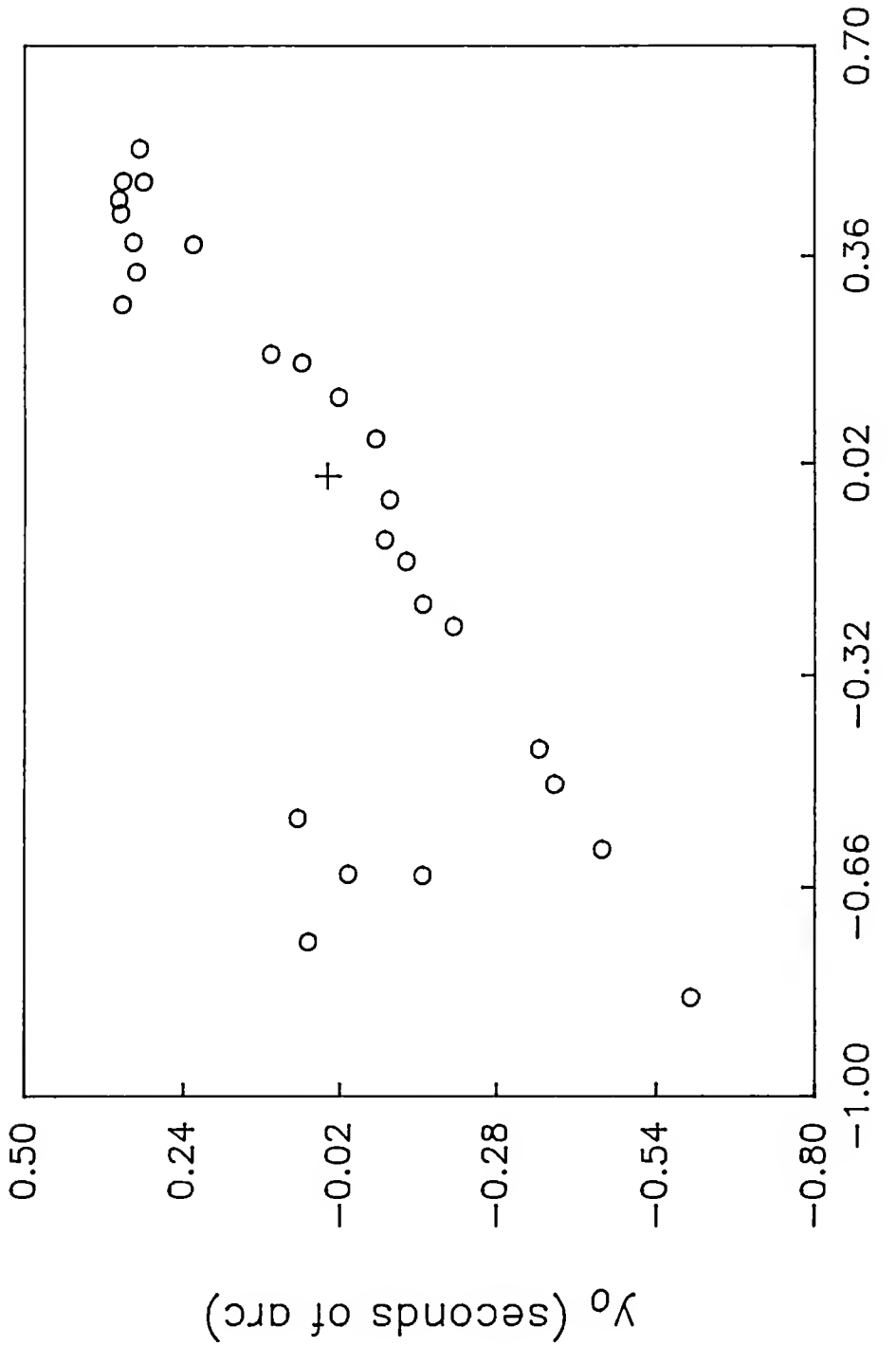

Figure 5-1. Plot of the observation data for β738 in the x_0-y_0 plane.

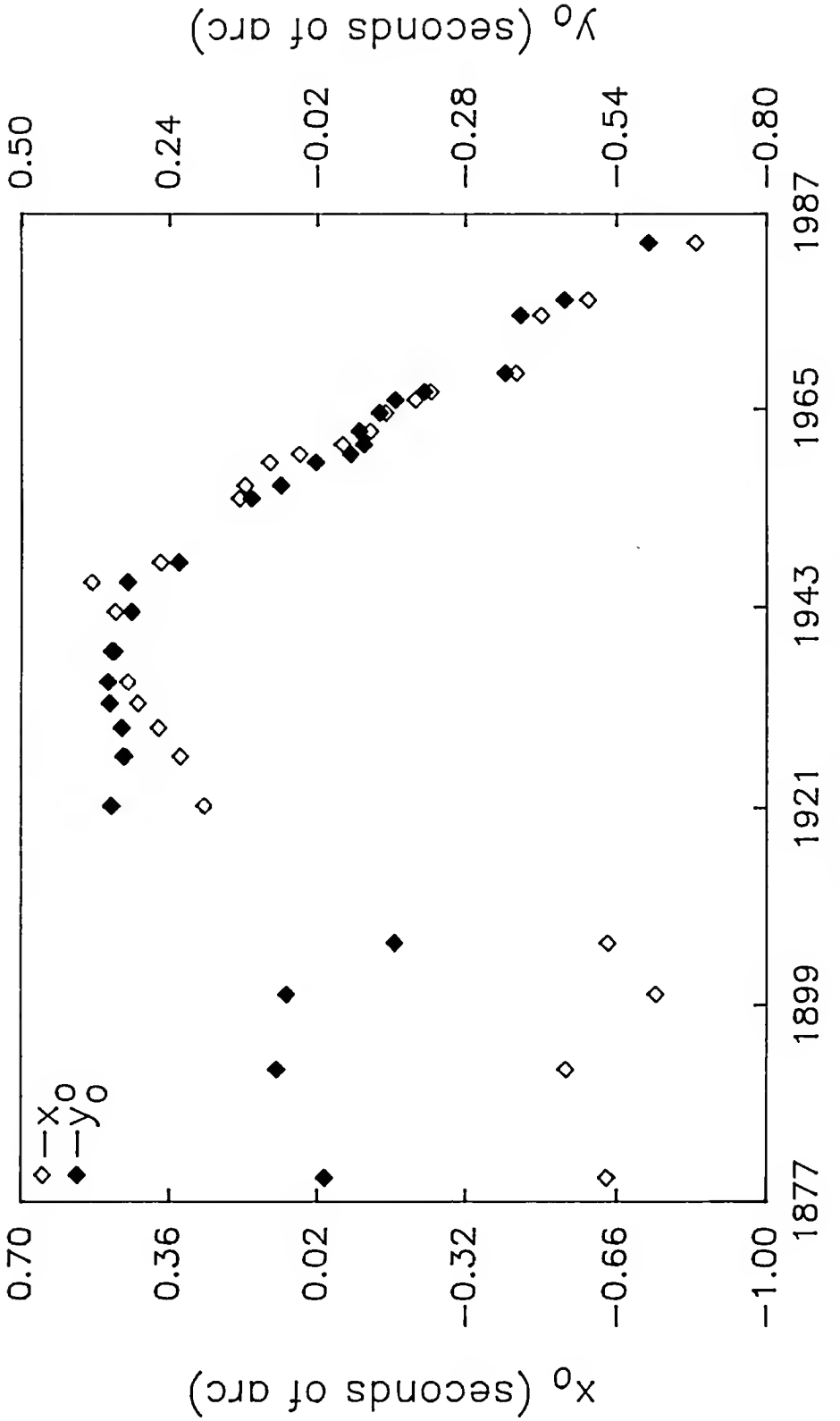

Figure 5-2. Plot of a) the x_0- b) y_0-coordinates against the observing epochs of the observation data for β738.

Table 5-3.

The Initial Approximate Solution $\hat{\mathbf{a}}_0$ for β738

P_0	T_0	a_0''	e_o	i_0°	ω_0°	Ω_0°
156.65	1955.2	0.958	0.523	105.2	57.4	31.2

Table 5-4.

The Solution #1 for β378

	P(yrs)	T	a"	e	i°	ω°	Ω°
solution #1	106.5	1952.9	7.426	0.241	101.6	48.6	26.1
standard deviations	4.73	2.09	0.038	0.033	1.55	8.78	2.37
standard deviations of the uncorrelated parameters	0.023	20.38	0.021	0.012	16.70	4.93	2.32
the efficiency				0.361			
the covariance matrix	0.00105	0.00001	-0.00025	-0.00049	0.00004	0.00128	-0.00009
	0.00001	0.00000	0.00000	0.00001	0.00000	0.00004	0.00000
	-0.00025	0.00000	0.00035	-0.00014	-0.00008	-0.00016	0.00046
	-0.00049	0.00001	0.00014	0.00084	-0.00001	0.00057	0.00006
	0.00004	0.00000	-0.00008	-0.00001	0.00005	0.00009	-0.00015
	0.00128	0.00004	-0.00016	0.00057	0.00009	0.00527	0.00015

Table 5-4 (continued)

	P(yrs)	T	a"	e	i°	ω°	Ω°
the correlation matrix	-1.00000	0.46336	-0.41513	-0.52158	0.18247	-0.54280	-0.07502
	0.46336	1.00000	-0.13857	0.38868	0.20111	0.97796	-0.02163
	-0.41513	-0.13857	1.00000	0.26090	-0.60222	-0.11928	0.67952
	-0.52158	0.38868	0.26090	1.00000	-0.05161	0.27167	0.05509
	0.18247	0.20111	-0.60222	-0.05161	1.00000	0.17910	-0.60616
	0.54280	0.97796	-0.11928	0.27167	0.17910	1.00000	0.05660
	-0.07502	-0.02163	0.67952	0.05509	-0.60616	0.05660	1.00000
the transformation matrix	-0.00102	0.99997	0.00011	-0.00320	0.00120	-0.00629	0.00120
	-0.08818	0.00125	-0.16364	-0.06186	-0.97745	0.04201	-0.06672
	0.59839	0.00132	0.59638	0.42977	-0.17410	-0.16814	-0.20730
	0.37053	0.00059	-0.68660	0.51859	0.02054	-0.18073	0.29877
	-0.25804	-0.00654	0.03721	-0.08680	-0.01713	-0.96103	-0.02395
	0.43660	-0.00065	-0.36523	-0.42745	0.09272	-0.07716	-0.69189
	0.48954	-0.00262	0.10657	-0.59353	-0.07052	-0.08788	0.61969

Table 5-5.

Residuals of the Observations in θ,ρ and x,y for Solution #1 of β738

	t	v_θ	v_ρ	v_x	v_y
1	1879.70	22.281	0.201	-0.121	-0.326
2	1891.80	24.309	0.229	-0.189	-0.305
3	1900.21	15.235	-0.138	0.153	-0.168
4	1905.98	-8.507	-0.203	0.186	0.115
5	1921.25	30.522	-0.239	-0.248	-0.141
6	1926.67	7.713	-0.121	-0.122	-0.053
7	1929.77	3.396	-0.076	-0.075	-0.030
8	1932.45	0.244	-0.059	-0.047	-0.035
9	1934.76	-1.703	-0.030	-0.014	-0.031
10	1938.16	-3.224	0.002	0.020	-0.026
11	1942.52	-4.821	0.041	0.060	-0.023
12	1845.80	-5.931	-0.038	-0.006	-0.073
13	1947.98	-8.737	0.100	0.122	-0.023
14	1955.00	-19.091	0.075	0.095	-0.062
15	1956.45	345.659	0.043	0.048	-0.048
16	1959.03	-23.659	0.007	-0.012	-0.054
17	1959.97	0.265	0.020	0.013	-0.016
18	1961.03	31.476	0.014	0.062	-0.019
19	1962.53	30.273	0.024	0.055	-0.062
20	1964.58	10.992	0.058	-0.006	-0.071
21	1966.02	10.225	0.051	-0.003	-0.072
22	1966.94	3.779	0.032	-0.008	-0.038
23	1969.04	0.972	-0.120	0.098	0.069
24	1975.55	-5.478	0.046	-0.073	0.024
25	1977.28	-6.502	-0.039	-0.014	0.091
26	1983.71	-8.721	-0.208	0.105	0.228

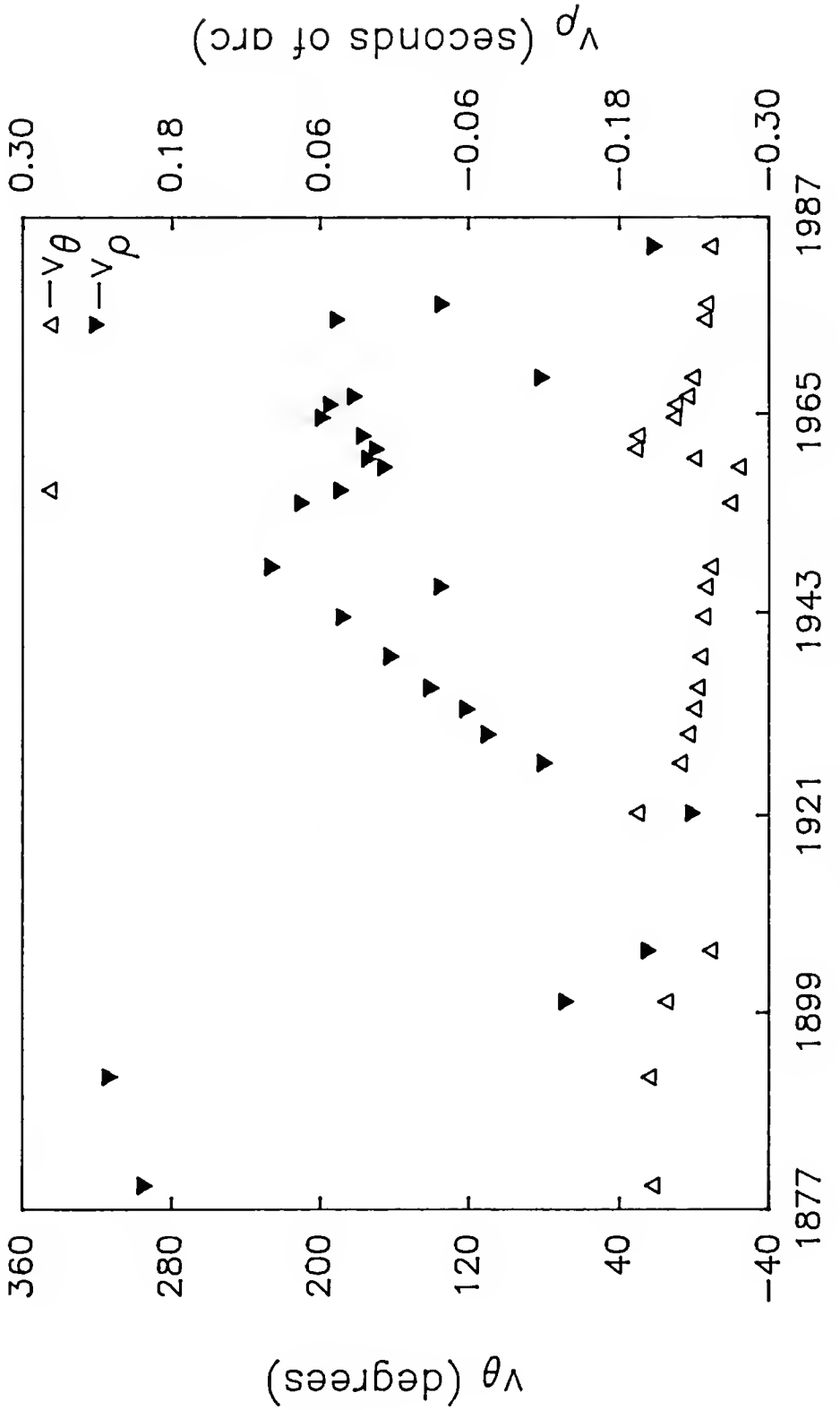

Figure 5-3. The residuals of the observations for β738 in (ρ,θ) according to solution #1.

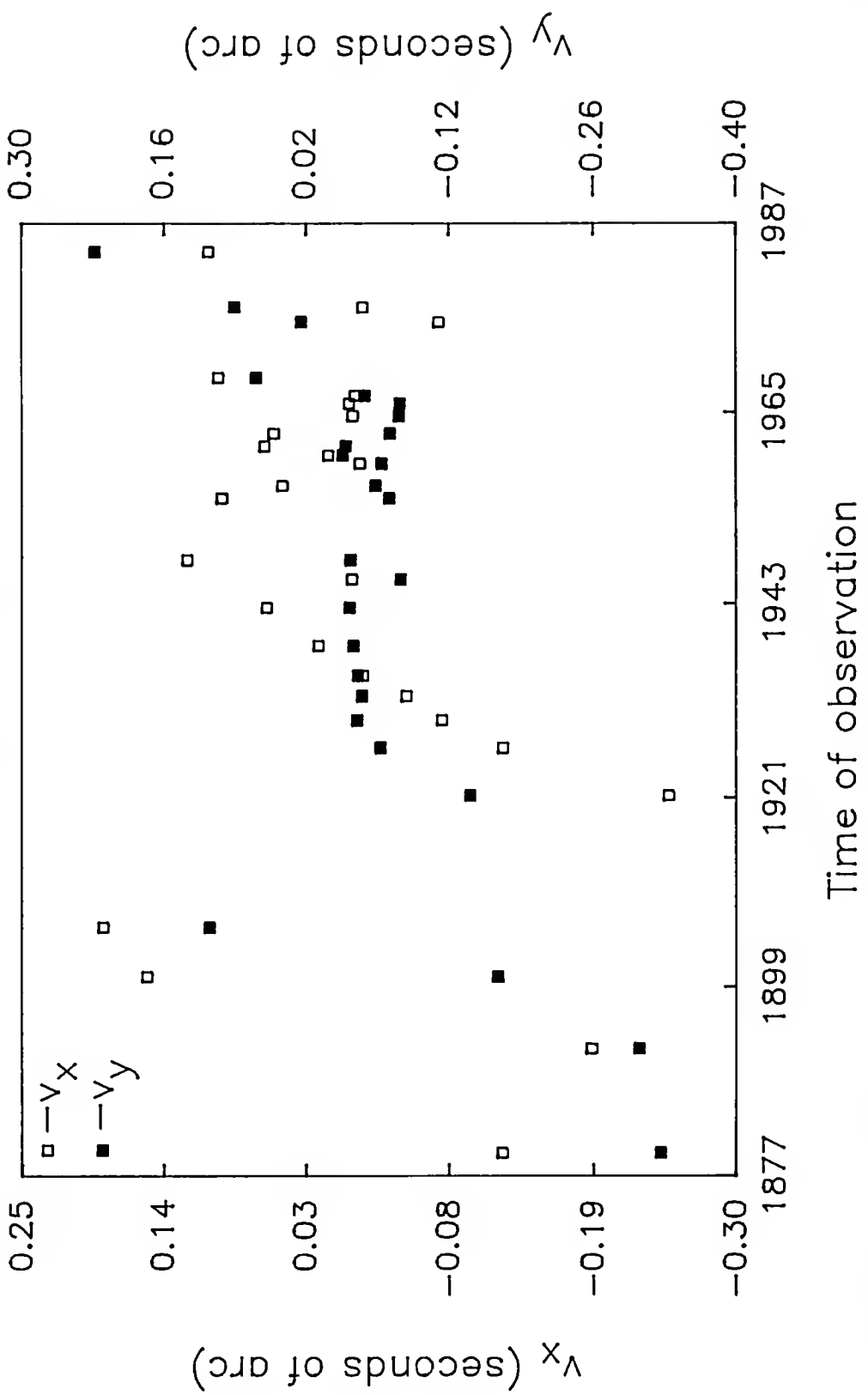

Figure 5-4. The residuals of the observations for β738 in (x,y) according to solution '#1.

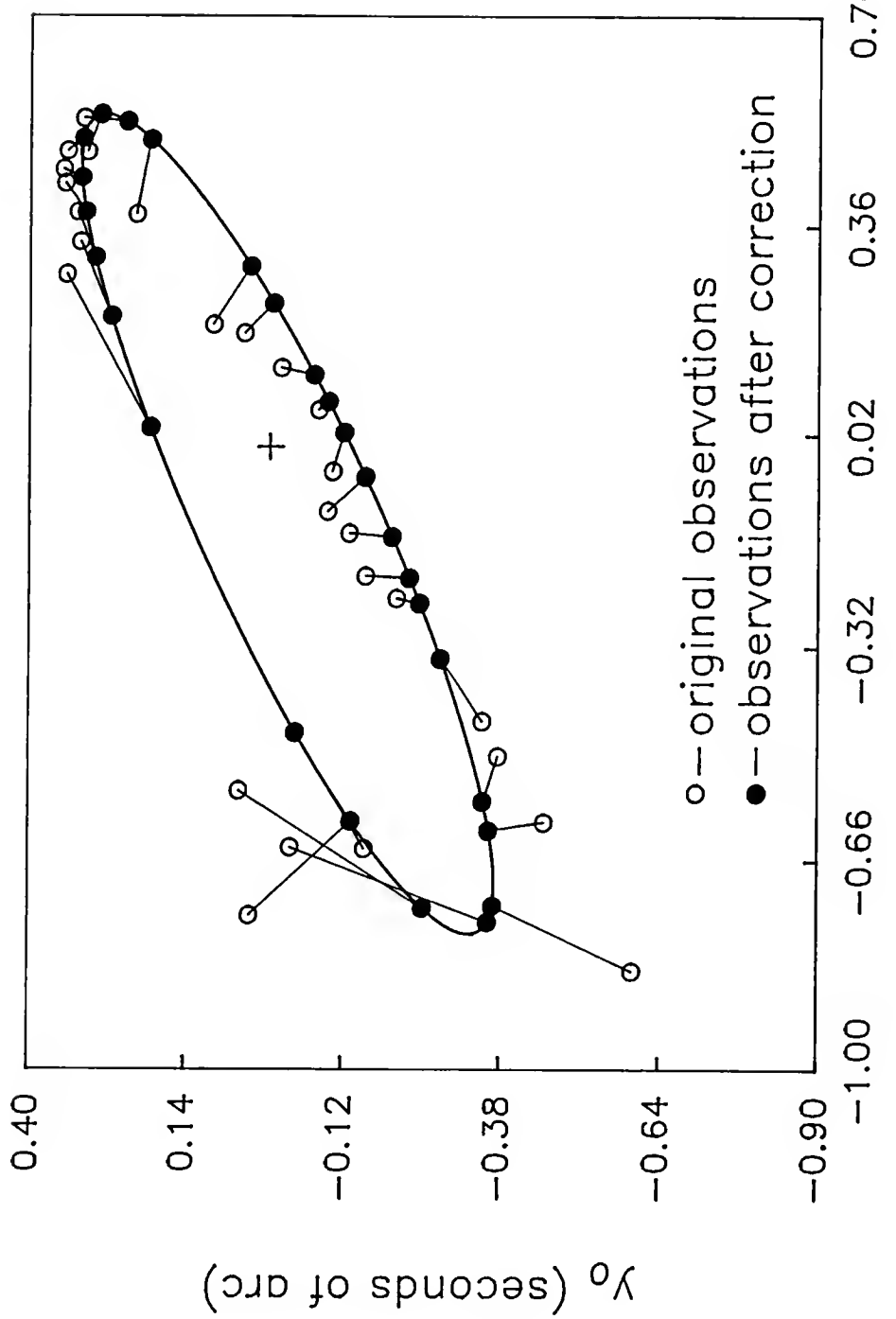

Figure 5-5. The original observations for β738 compared with the observations after correction according to solution #1.

Table 5-6.

Heintz' Result for β738

P(yr)	T	a"	e	i°	ω°	Ω°
290.0	1952.0	1.460	0.63	96.7	38.0	29.4

Table 5-7.

The Solution #2 for β378

	P(yrs)	T	a"	e	i°	ω°	Ω°
solution #2	291.95	1953.5	1.281	0.630	96.59	47.05	31.37
standard deviations	90.90	1.31	0.377	0.109	0.78	7.02	1.07
standard deviations of the uncorrelated parameters	0.009	7.29	0.009	0.006	57.82	1.81	2.56
the efficiency				0.064			
the covariance matrix	0.06440	0.00010	-0.09938	-0.06871	0.00031	0.02444	0.00482
	0.00010	0.00000	-0.00017	-0.00010	0.00000	0.00008	0.00000
	-0.09938	-0.00017	0.15552	0.10539	-0.00072	-0.04038	-0.00676
	-0.06871	-0.00010	0.10539	0.07398	-0.00023	-0.02446	-0.00529
	0.00031	0.00000	-0.00072	-0.00023	0.00006	0.00047	-0.00012
	0.02444	0.00008	-0.04038	-0.02446	0.00047	0.01496	0.00141
	0.00482	0.00000	-0.00676	-0.00529	-0.00012	0.00141	0.00118

Table 5-7 (continued)

	P(yrs)	T	a"	e	i°	ω°	Ω°
the correlation matrix	1.00000	0.59218	-0.99306	-0.99550	0.16324	0.78747	0.55320
	0.59218	1.00000	-0.13857	-0.65664	-0.52017	0.95382	0.17842
	-0.99306	-0.65664	1.00000	0.98252	-0.24660	-0.83706	-0.49879
	-0.99550	-0.52017	0.98252	1.00000	-0.11618	-0.73509	-0.56656
	0.16324	0.59504	-0.24660	-0.11618	1.00000	0.52217	-0.45293
	0.78747	0.95382	-0.83706	-0.73509	0.52217	1.00000	0.33507
	0.55320	0.17842	-0.49879	-0.56656	-0.45293	0.33507	1.00000
the transformation matrix	-0.00602	0.99995	0.00014	-0.00586	0.00163	-0.00549	0.00071
	-0.09580	0.00107	-0.00281	-0.07470	-0.97974	0.07131	-0.14234
	0.74444	0.00742	0.00185	0.62960	-0.13253	-0.17810	-0.00842
	0.40571	0.00243	0.58166	-0.37333	0.08224	0.34384	-0.48239
	-0.46068	-0.00077	0.71651	0.49012	-0.00283	-0.18190	-0.03295
	-0.18105	-0.00157	-0.32895	0.14033	0.11015	-0.30565	-0.85660
	0.16429	-0.00623	0.20018	-0.44577	-0.06034	-0.84760	0.11007

Table 5-8.

Residuals in ρ,θ,x and y of the Observations for $\beta 738$ in Solution #2

	t	v_θ	v_ρ	v_x	v_y
1	1879.70	2.160	-0.207	0.208	-0.005
2	1891.80	-10.692	-0.309	0.316	0.016
3	1900.21	-54.998	-0.585	0.661	0.106
4	1905.98	-107.227	-0.473	0.652	0.344
5	1921.25	-2.623	-0.039	-0.011	-0.042
6	1926.67	-0.680	0.020	0.018	0.009
7	1929.77	0.579	0.019	0.011	0.016
8	1932.45	0.384	-0.002	-0.004	0.002
9	1934.76	0.248	-0.002	-0.003	0.001
10	1938.16	0.715	-0.004	-0.007	0.004
11	1942.52	1.131	0.011	0.003	0.015
12	1845.80	1.485	-0.073	-0.070	-0.024
13	1947.98	-0.228	0.066	0.058	0.032
14	1955.00	-4.324	0.043	0.046	0.000
15	1956.45	3.020	0.006	0.004	0.011
16	1959.03	-5.384	-0.047	-0.048	-0.001
17	1959.97	4.197	-0.039	-0.020	0.034
18	1961.03	16.641	-0.034	0.034	0.027
19	1962.53	16.383	-0.003	0.032	-0.023
20	1964.58	4.004	0.044	-0.020	-0.041
21	1966.02	6.319	0.041	-0.011	-0.050
22	1966.94	1.320	0.024	-0.013	-0.022
23	1969.04	0.912	-0.122	0.099	0.071
24	1975.55	-1.468	0.079	-0.073	-0.033
25	1977.28	-1.755	0.010	-0.022	0.013
26	1983.71	-1.629	-0.068	0.040	0.062

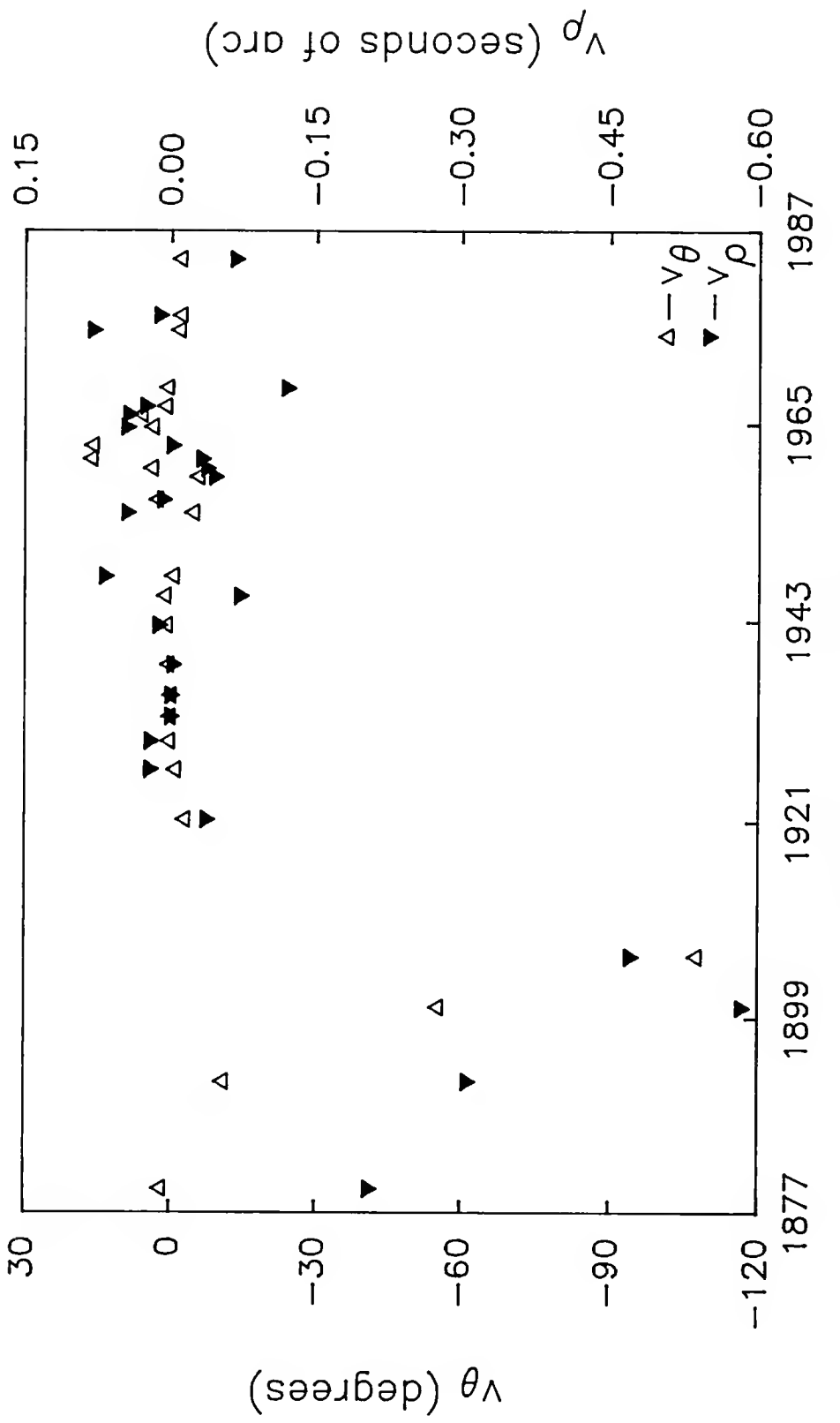

Figure 5-6. The residuals of the observations for β738 in (ρ,θ) according to solution #2.

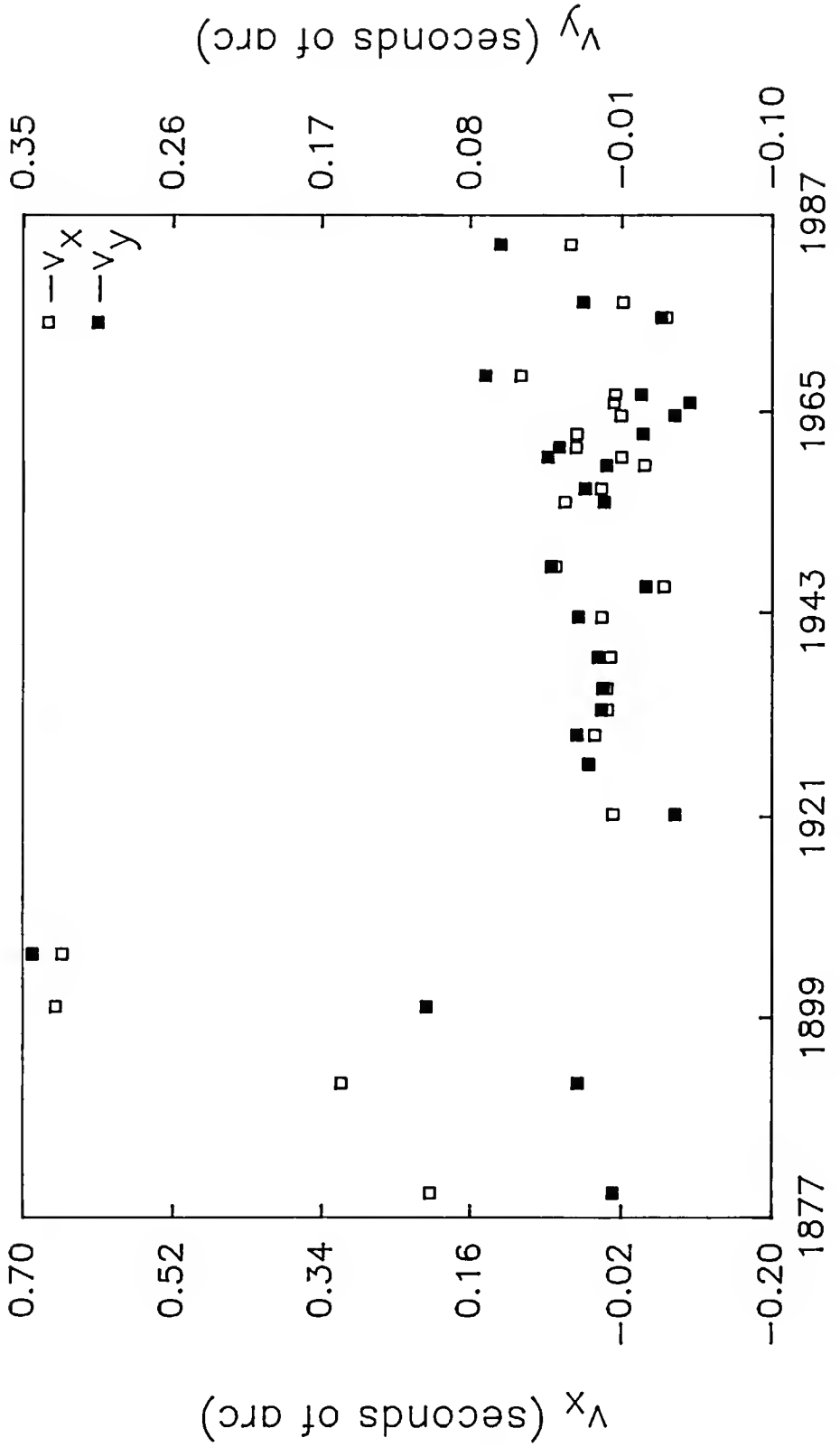

Figure 5-7. The residuals of the observations for β738 in (x,y) according to solution #2.

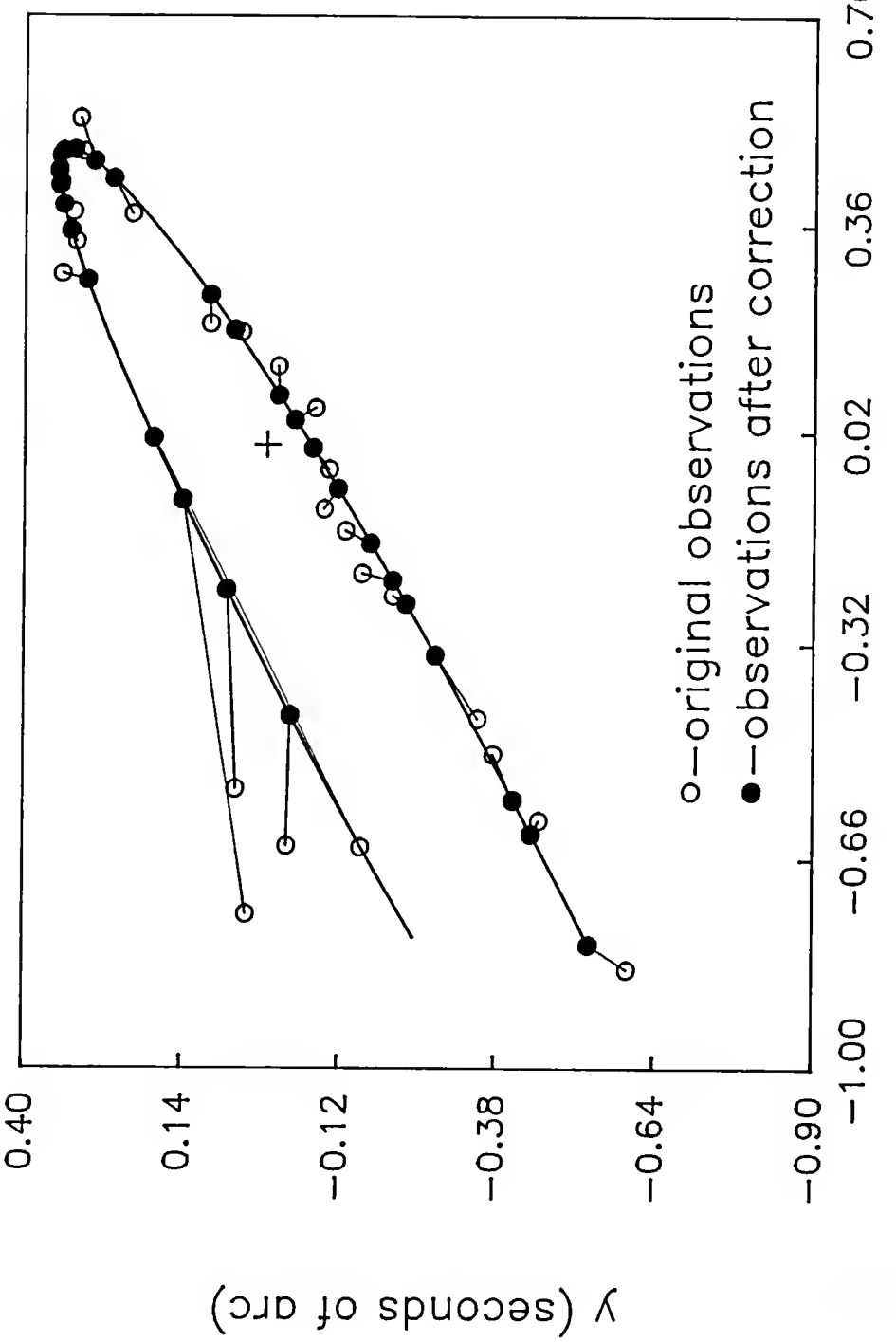

Figure 5-8. The original observations for β738 compared with the observations after correction according to solution #2.

Table 5-9.

The Observation Data for BD+19°5116
(courtesy W. D. Heintz)

+19°5116	23319	N1956	(2000)	10.4-12.1 M_4Ve
1945.473	174.47	3.563	2n	11ex S
49.809	165.72	3.430	3	21
51.964	161.08	3.435	3	14
54.804	154.45	3.583	2	10
55.971	152.54	3.565	2	10
58.482	147.18	3.616	3	14
60.152	143.91	3.648	6	50
61.682	141.51	3.708	4	46
62.835	139.46	3.759	8	70
63.757	138.49	3.770	4	31
64.860	136.50	3.810	6	56
65.766	134.67	3.840	3	23
68.289	129.65	3.967	2	21
69.838	127.34	4.101	6	76
70.792	126.12	4.098	2	28
71.939	124.54	4.134	5	36
73.903	121.48	4.236	7	32
75.872	118.90	4.297	2	16
77.049	117.65	4.361	4	25
78.904	115.01	4.431	2	12
80.529	112.47	4.621	3	12
81.856	111.72	4.602	4	18
1983.581	108.33	4.676	3	19
1941.75	182.17	3.678	9	M
42.65	179.70	3.643	6	
43.79	177.20	3.750	5	
44.72	174.82	3.679	4	
46.03	172.59	3.605	5	
49.75	165.33	3.644	4	
1955.78	153.38	3.605	5	
1952.19	161.8	3.52	7	VB
59.39	145.2	3.82	6	C Wor 3
62.07	141.1	3.76	11	B VB 4 C 3
65.53	137.0	3.97	7	VB
1981.64	112.3	4.24	2	hz

Pg. positions from parallax plates at Swarthmore (S) refractor (meas. Heintz), from McCormick (M) refractor (meas. Eichhorn); few micrometer observations. Position angles equator 2000.

Table 5-10.

The Reduced Initial Data for BD+19°5116

	t	θ_0	ρ_0	x_0	y_0
1	1941.750	182.17	3.678	-3.6754	-0.1393
2	1942.650	179.70	3.643	-3.6430	0.0191
3	1943.790	177.20	3.750	-3.7455	0.1832
4	1944.720	174.82	3.679	-3.6640	0.3322
5	1945.473	174.47	3.563	-3.6454	0.3434
6	1946.030	172.59	3.605	-3.5749	0.4649
7	1949.750	165.33	3.644	-3.5252	0.9228
8	1949.809	165.72	3.430	-3.3240	0.8460
9	1951.964	161.08	3.435	-3.2494	1.1138
10	1952.190	161.80	3.520	-3.3439	1.0994
11	1954.804	154.45	3.583	-3.2326	1.5453
12	1955.780	153.38	3.605	-3.2229	1.6153
13	1955.971	152.54	3.565	-3.1633	1.6439
14	1958.482	147.18	3.616	-3.0388	1.9599
15	1959.390	145.20	3.820	-3.1368	2.1801
16	1960.152	143.91	3.648	-2.9479	2.1489
17	1961.682	141.51	3.708	-2.9023	2.3078
18	1962.070	141.10	3.760	-2.9262	2.3611
19	1962.838	139.46	3.759	-2.8567	2.4433
20	1963.757	138.49	3.770	-2.8231	2.4986
21	1964.860	136.50	3.810	-2.7637	2.6226
22	1965.530	137.00	3.970	-2.9035	2.7075
23	1965.766	134.67	3.840	-2.6996	2.7309
24	1968.289	129.65	3.967	-2.5313	3.0544
25	1969.838	127.34	4.101	-2.4874	3.2605
26	1970.792	126.12	4.098	-2.4157	3.3103
27	1971.939	124.54	4.134	-2.3439	3.4053
28	1973.903	121.48	4.236	-2.2120	3.6126
29	1975.872	118.90	4.297	-2.0767	3.7619
30	1977.049	117.65	4.361	-2.0238	3.8630
31	1978.904	115.01	4.431	-1.8733	4.0155
32	1980.529	112.47	4.621	-1.7661	4.2707
33	1981.640	112.30	4.240	-1.6089	3.9229
34	1981.856	111.72	4.602	-1.7031	4.2753
35	1983.581	108.33	4.676	-1.4706	4.4387

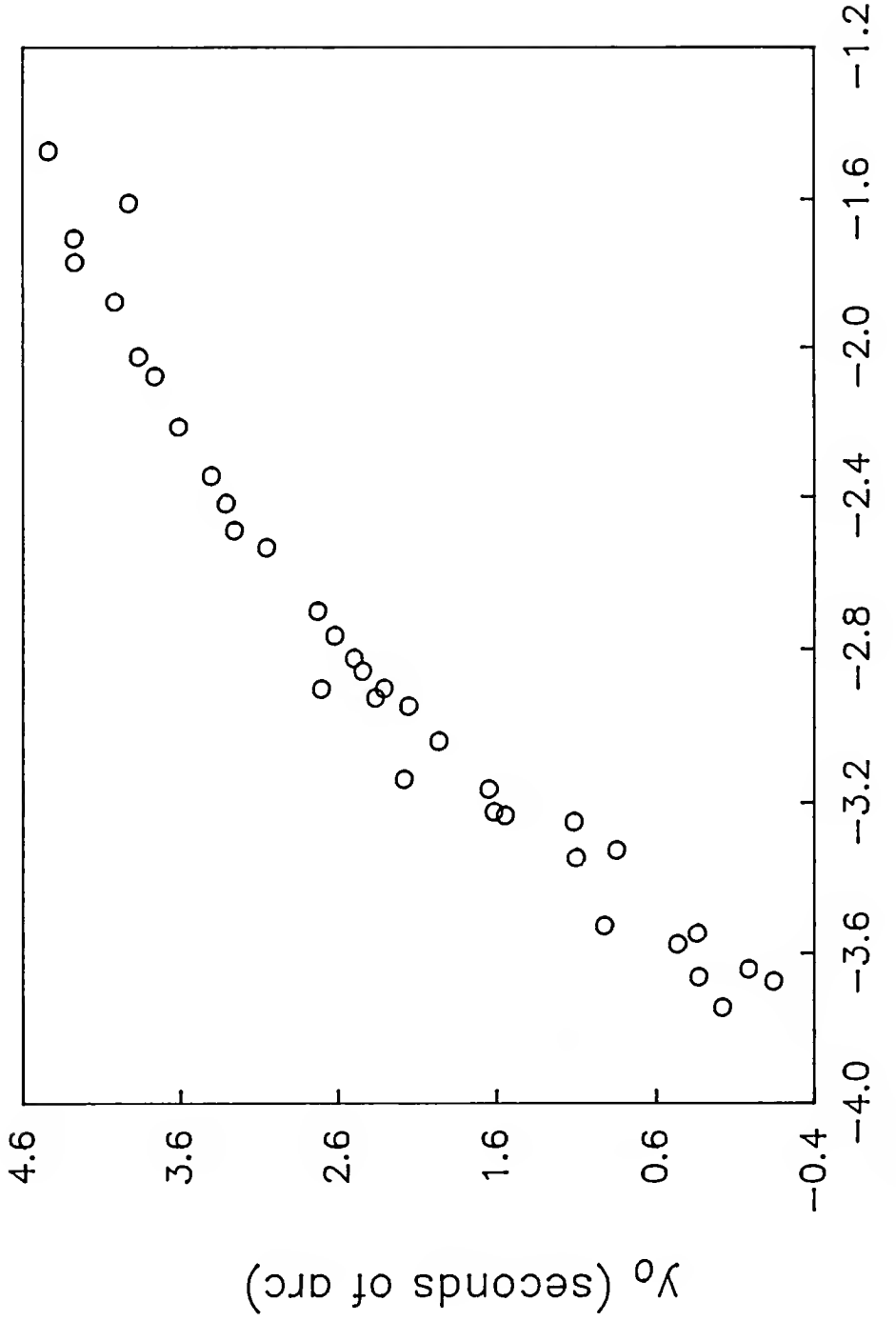

Figure 5-9. Plot of the observation data for BD+19°5116 in the x_0-y_0 plane.

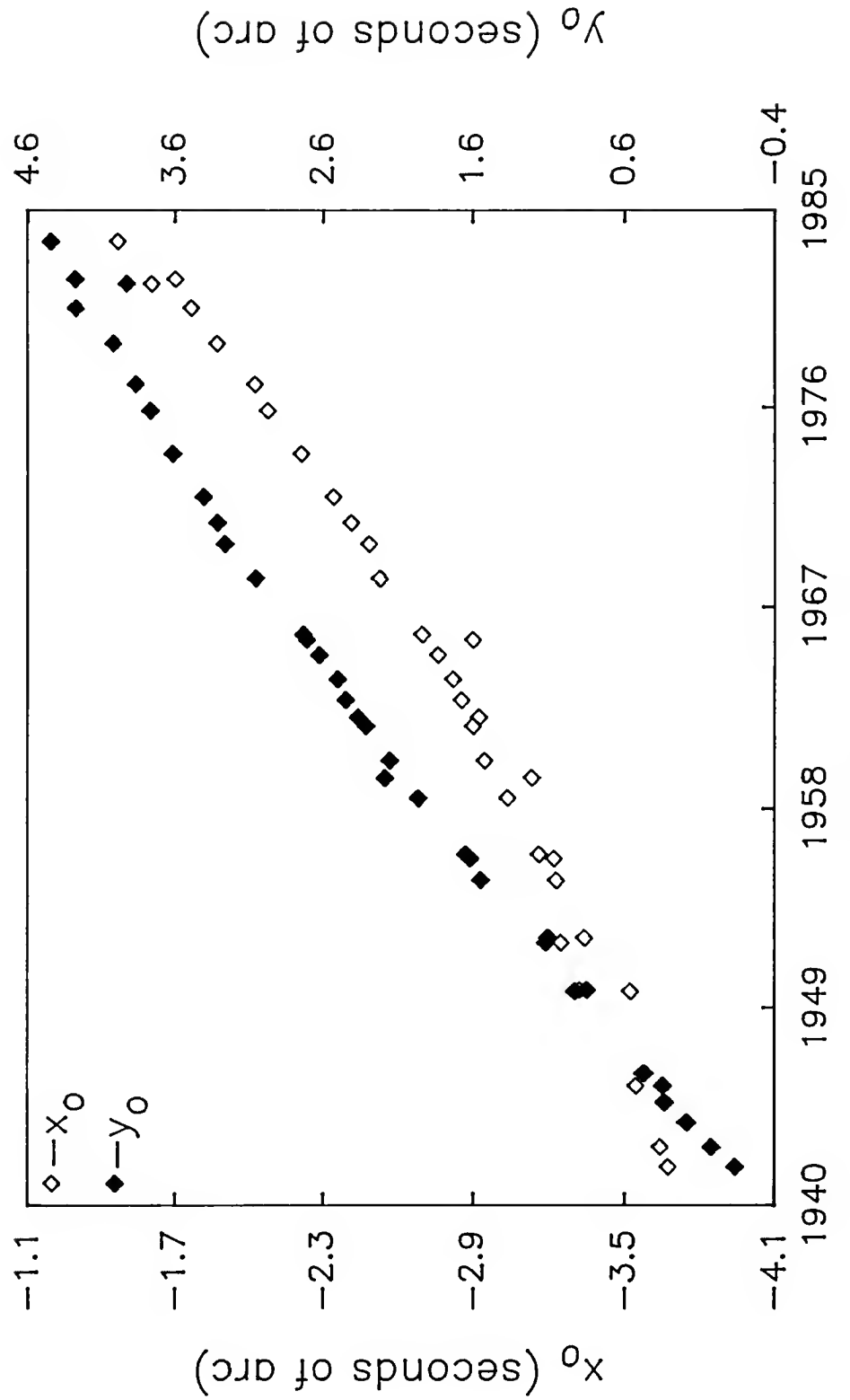

Figure 5-10. Plot of a) the x_0- b) y_0-coordinates against the observing epochs of the observations for BD+19°5116.

Table 5-11.

The Initial Approximate Solution $\hat{a}_0$ for BD+19°5116

P_0(yrs)	T_0	a_0''	e_0	i_0°	ω_0°	Ω_0°
125.6	1891.1	5.31	0.61	115.2	94.0	76.0

Table 5-12.

The Final Solution for BD+19°5116 by MQ Method

	P(yrs)	T	a"	e	i°	ω°	Ω°
the final solution	130.88	1891.10	5.733	0.658	112.24	93.36	70.28
standard deviations	98.70	238.68	17.16	0.148	56.64	9.02	41.36
standard deviations of the uncorrelated parameters	0.0044	1.238	0.028	0.002	5.42	63.22	14.32
the efficiency				0.00018			
the covariance matrix	12.86588	0.55060	14.11423	30.59438	-4.41831	-0.18458	-2.30953
	0.55060	0.02366	0.60254	1.30807	-0.18866	-0.00689	-0.10037
	14.11423	0.60254	15.51671	33.59297	-4.85616	-0.21927	-2.51029
	30.59438	1.30807	33.59297	72.77996	-10.51481	-0.45297	-5.47306
	-4.41831	-0.18866	-4.85616	-10.51481	1.51984	0.06812	0.78650
	-0.18458	-0.00689	-0.21927	-0.45297	0.06812	0.01366	0.01665
	-2.30953	-0.10037	-2.51029	-5.47306	0.78650	0.01665	0.43986

Table 5-12 (continued)

	P(yrs)	T	a"	e	i°	ω°	Ω°
the correlation matrix	1.00000	0.99818	0.99907	0.99984	-0.99928	-0.45695	-0.97320
	0.99818	1.00000	0.99507	0.99718	-0.99548	-0.40265	-0.98525
	0.99907	0.99507	1.00000	0.99968	-0.99999	-0.49011	-0.96431
	0.99983	0.99718	0.99968	1.00000	-0.99979	-0.46990	-0.97003
	-0.99928	-0.99548	-0.99999	-0.99979	1.00000	0.48690	0.96524
	-0.45695	-0.40265	-0.49011	-0.46990	0.48690	1.00000	0.24247
	-0.97320	-0.98525	-0.96431	-0.97003	0.96524	0.24247	1.00000
the transformation matrix	0.03975	-0.99634	-0.00037	-0.00443	-0.03125	0.06507	-0.02247
	0.01892	-0.03091	0.21397	0.03665	0.97512	-0.00385	0.02610
	0.59874	0.02313	0.52455	-0.52653	-0.10091	-0.16982	-0.22264
	-0.09483	-0.05175	-0.22637	0.11484	0.05393	-0.85351	-0.43877
	-0.65558	-0.00198	0.55415	-0.02590	-0.09456	0.16166	-0.47684
	-0.27365	-0.05398	0.41417	-0.03833	-0.10707	-0.46074	0.72488
	-0.35483	-0.01514	-0.38626	-0.84028	0.12110	0.00528	0.06243

Table 5-13.

Residuals of the Observations for BD+19°5116
in θ, ρ, x and y

	t	v_θ	v_ρ	v_x	v_y
1	1941.750	-1.3854	-0.0141	0.0119	0.0891
2	1942.650	-0.6496	0.0041	-0.0037	0.0414
3	1943.790	-0.3677	-0.1215	0.1226	0.0173
4	1944.720	0.1869	-0.0636	0.0623	-0.0175
5	1945.473	-0.9591	0.0433	-0.0368	0.0636
6	1946.030	-0.1746	-0.0046	0.0060	0.0103
7	1949.750	-0.3556	-0.0633	0.0670	0.0054
8	1949.809	-0.8641	0.1506	-0.1323	0.0894
9	1951.964	-0.5481	0.1505	-0.1311	0.0812
10	1952.190	-1.7206	0.0667	-0.0282	0.1226
11	1954.804	0.4275	0.0270	-0.0359	-0.0127
12	1955.780	-0.4239	0.0180	-0.0040	0.0320
13	1955.971	0.0417	0.0608	-0.0552	0.0257
14	1958.482	0.5414	0.0549	-0.0648	0.0005
15	1959.39	0.7957	-0.1293	0.0772	-0.1161
16	1960.152	0.6524	0.0607	-0.0737	0.0015
17	1961.682	0.2183	0.0404	-0.0405	0.0140
18	1962.070	-0.0807	-0.0008	0.0039	0.0036
19	1962.835	0.1737	0.0224	-0.0244	0.0058
20	1963.757	-0.5043	0.0394	-0.0072	0.0511
21	1964.860	-0.4534	0.0348	-0.0042	0.0459
22	1965.530	-2.1135	-0.1028	0.1743	0.0324
23	1965.766	-0.1890	0.0353	-0.0157	0.0341
24	1968.289	0.6033	-0.0009	-0.0314	-0.0275
25	1969.838	0.4151	-0.0756	0.0227	-0.0779
26	1970.792	0.1332	-0.0350	0.0130	-0.0339
27	1971.939	-0.0557	-0.0250	0.0174	-0.0183
28	1973.903	0.0669	-0.0465	0.0202	-0.0422
29	1975.872	-0.1861	-0.0257	0.0246	-0.0158
30	1977.049	-0.5783	-0.0409	0 0577	-0.0161
31	1978.904	-0.4527	-0.0345	0.0461	-0.0167
32	1980.529	-0.0453	-0.1591	0.0641	-0.1457
33	1981.640	-1.2982	0.2654	-0.0058	0.2832
34	1981.856	-0.9916	-0.0883	0.1055	-0.0537
35	1983.581	0.2496	-0.0976	0.0118	-0.0990

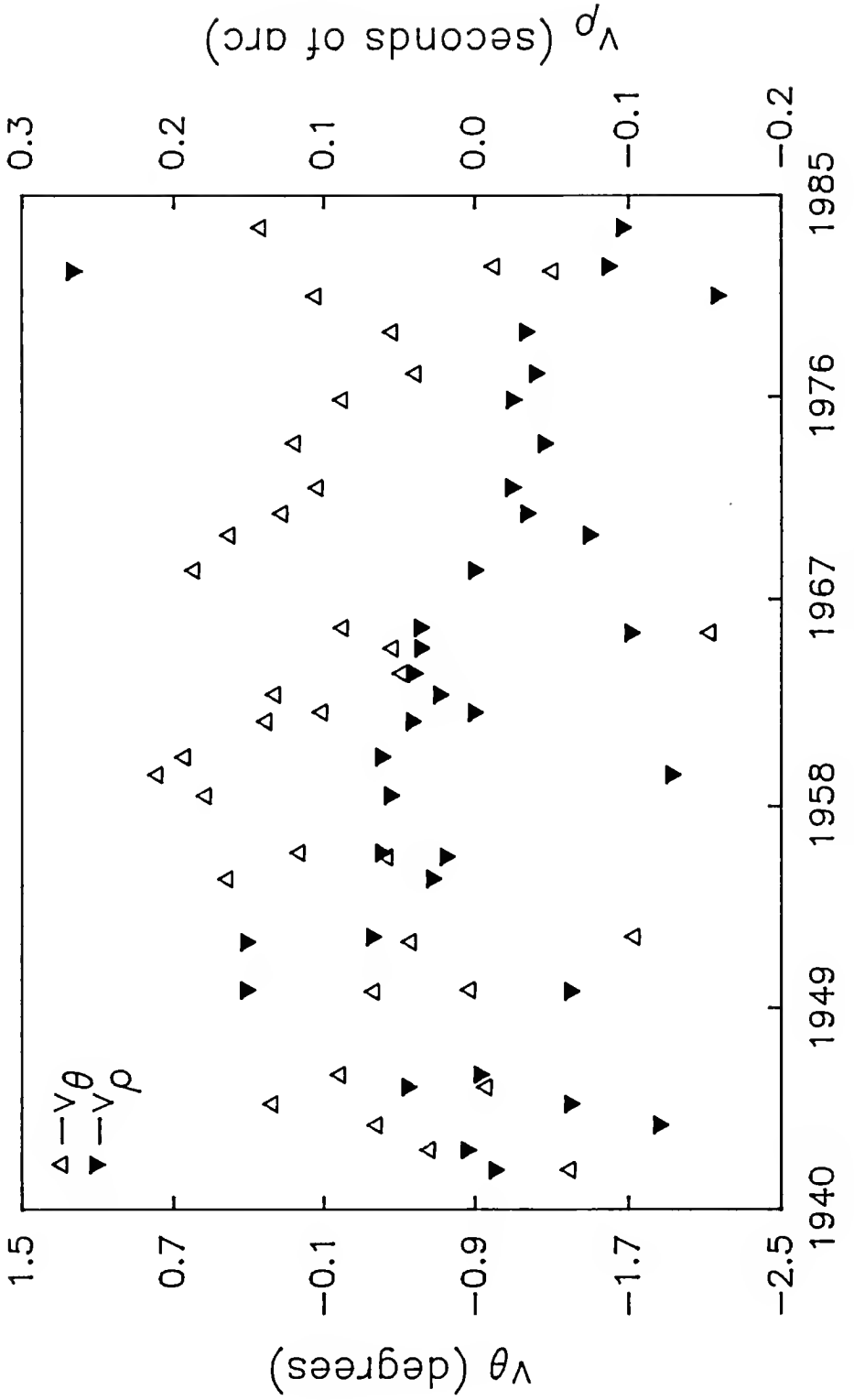

Figure 5-11. The residuals of the observations for BD+19°5116 in (ρ,θ).

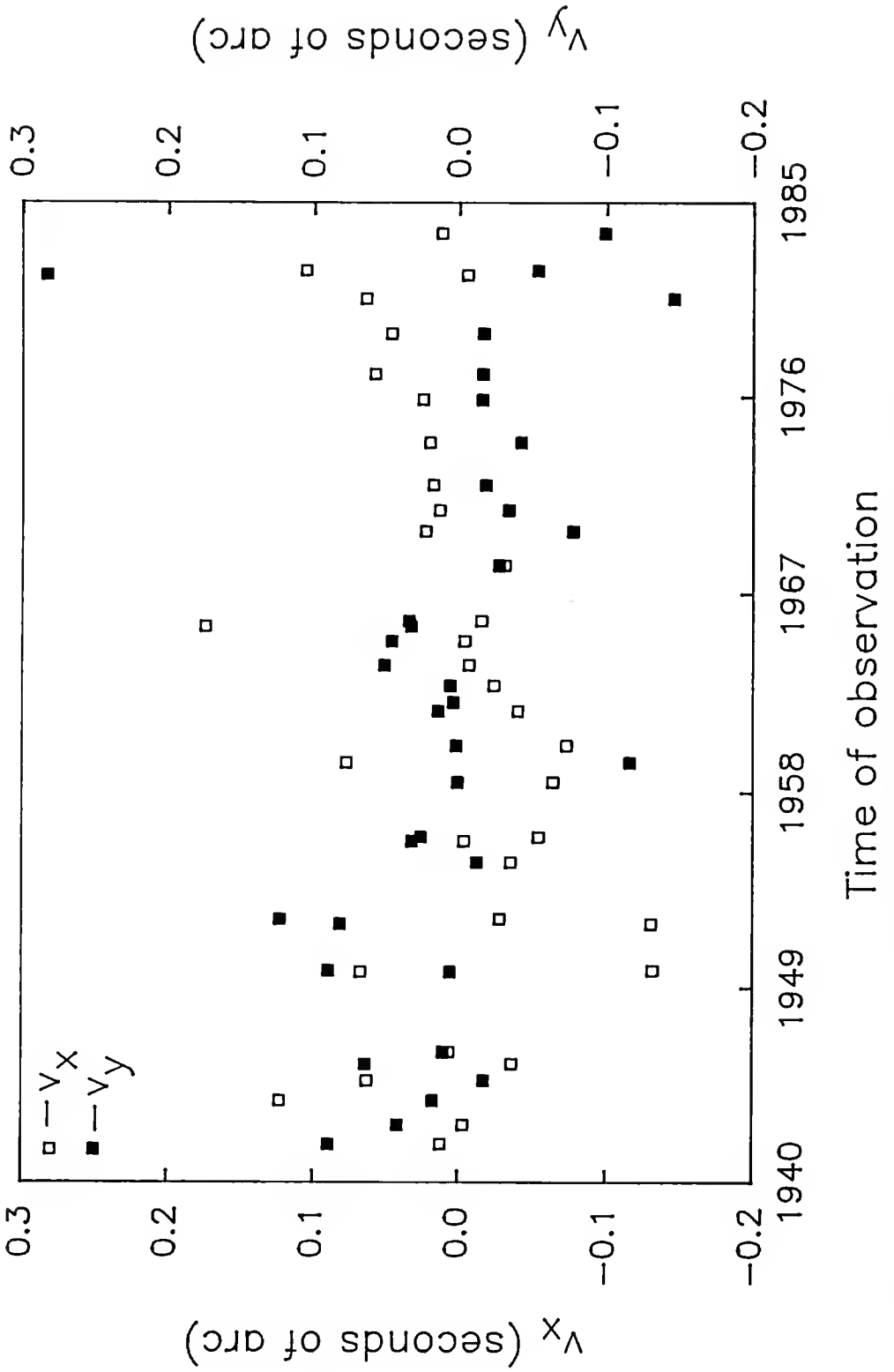

Figure 5-12. The residuals of the observations for BD+19°5116 in (x,y).

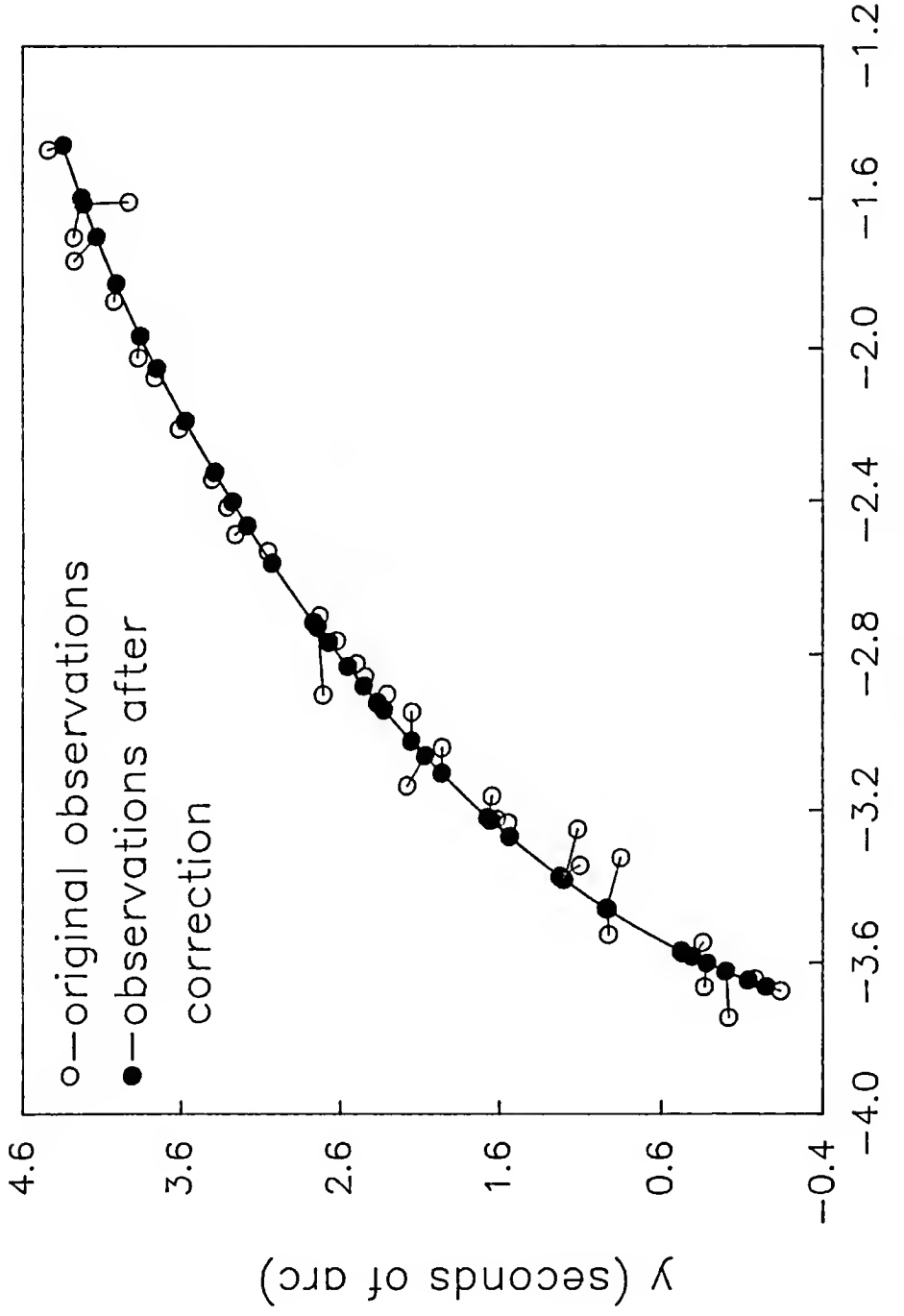

Figure 5-13. The original observations for BD+19° 5116 compared with the observations after correction.

CHAPTER VI
DISCUSSION

For solving our orbit problem, the Newton scheme and its modifications ("FP" and "MQ" algorithms) have been discussed in detail. As we have seen, the Newton scheme is powerful and converges very fast if the observational errors are sufficiently small and the initial approximations to the parameters are sufficiently accurate. But in other cases, Newton's scheme fails to converge. Some modifications of it were carried out for these cases. Two other schemes, "FP" and "MQ", are proposed. These two converge to a definitive solution even when Newton's scheme diverges. As we have seen, "FP" method only shortens, however, the length of the step and retains all other features of Newton's scheme, so that sometimes it converges still very slow. The "MQ" method seems more powerful because it not only shortens the length of the step but also makes the corrections to the parameters gradually uncorrelated in subsequent iterations. "MQ" method converges much faster than "FP" does.

In addition to which scheme is used, two things are important about the observation data themselves.

1) The sufficiency of observation data itself is of importance. Whether a set of observations is suitable for computing an orbit depends on the amount, consistency and distribution of the data. Some observation data define only a short arc of small curvature variation like in our example for BD+19°5116. The orbit-computation from such data is relatively more difficult.

2) The weighting of observation is also of importance. The more a computation is based on more precise observations the better the chances are of practical success. If using weak data is unavoidable, low weights must be assigned to them, in accordance with standard practice in the field.

REFERENCES

Aitken, Robert G. 1935. The Binary Stars, Dover Publications, Inc., New York, 70.

Brown, D. C. 1955. Report No. 937, Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland.

Comstock, G. C. 1918. The Astronomical Journal, 31, 33.

Eichhorn, Heinrich. 1985. Astrophysics and Space Science, 110, 119.

Eichhorn, Heinrich, and Warren G. Clary. 1974. Mon. Not. R. Astr. Soc., 166, 425.

Eichhorn, Heinrich, and Carl S. Cole. 1985. Celestial Mechanics, 37, 263.

Fletcher, W., and M. J. D. Powell. 1963. Comput. J., 6, 163.

Glasenapp, S. 1889. Mon. Not. Roy. Astron. Soc., 49, 276.

Heintz, Wulff D. 1971. Double Stars, D. Reidel Publishing Company, London, England, 38.

Herschel, John. 1833. Memoirs R.A.S., 5, 171.

Jefferys, William H. 1980. The Astronomical Journal, 85, 177.

Jefferys, William H. 1981. The Astronomical Journal, 86, 149.

Lanson, C. L., and R. J. Harsson. 1974. Solving Least Squares Problems, Englewood Cliffs, N.J. (Prentice-Hall).

Marquardt, D. W. 1963. J. Soc. Ind. Appl. Math., 11, 431.

Russell, H. N. 1898. Astrophys. J., 19, 9.

Smart, W. M. 1930. Mon. Not. Roy. Astron. Soc., 90, 534.

Thiele, T. N. 1883. Astron. Nachr., 104, 245.

van den Bos, W. H. 1926. Union Obs. Circ., 2, 356.

van den Bos, W. H. 1932. Union Obs. Circ., 4, 223.

Zwiers, J. H. 1896. Astron. Nachr., 139, 369.

BIOGRAPHICAL SKETCH

Yu-lin Xu was born on October 20, 1945, in Jinyun County, Zhejiang, China. He received his elementary education in Jinyun and was graduated from the Jinhua First Middle School in 1962. He entered the Zhejiang University, Hangzhou, in September 1962 and received his diploma from the Department of Optical Instrument Engineering in July 1968.

He was an engineer in Hangzhou Camera Institute, Hangzhou, China, from September 1968 through June 1983. In July 1983, he entered the University of Florida to start working toward a Ph.D. in astronomy.

In addition, he is working on the experimentation and theory for the microwave scattering at the Space Astronomy Laboratory, the University Florida.

He received the degree of Master of Science and the degree of Doctor of Philosophy from the University of Florida in April 1988.

He married Yue-mei Li in January 1974, and they now have two daughters, Ying-xun and Hui-yi.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Heinrich K. Eichhorn, Chairman
Professor of Astronomy

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Kwan-Yu Chen, Co-Chairman
Professor of Astronomy

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Frank B. Wood
Professor of Astronomy

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Haywood C. Smith
Associate Professor of Astronomy

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Philip Bacon

Philip Bacon
Associate Professor of
Mathematics

This dissertation was submitted to the Graduate Faculty of the Department of Astronomy in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

April 1988

Dean, Graduate School